

Real and Complex Monotone Communication Games

Gesualdo Scutari¹, Francisco Facchinei², Jong-Shi Pang³, and Daniel P. Palomar⁴

E-mail: gesualdo@buffalo.edu, facchinei@dis.uniroma1.it, jspang@uiuc.edu, palomar@ust.hk.

¹COMSENS Research Center, Dpt. of Electrical Engineering, State University of New York at Buffalo, Buffalo, USA.

²Dpt. of Computer, Control, and Management Engineering Antonio Ruberti, Univ. of Rome "La Sapienza", Rome, Italy.

³Dpt. of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, USA.

⁴Dpt. of Electronic and Computer Eng., Hong Kong Univ. of Science and Technology, Hong Kong.*

Submitted to *IEEE Transactions on Information Theory*, October 8, 2012.

Abstract

Noncooperative game-theoretic tools have been increasingly used to study many important resource allocation problems in communications, networking, smart grids, and portfolio optimization. In this paper, we consider a general class of convex Nash Equilibrium Problems (NEPs), where each player aims to solve an arbitrary smooth convex optimization problem. Differently from most of current works, we do not assume any specific structure for the players' problems, and we allow the optimization variables of the players to be matrices in the complex domain. Our main contribution is the design of a novel class of distributed (asynchronous) best-response- algorithms suitable for solving the proposed NEPs, even in the presence of *multiple* solutions. The new methods, whose convergence analysis is based on Variational Inequality (VI) techniques, can select, among all the equilibria of a game, those that optimize a given performance criterion, at the cost of limited signaling among the players. This is a major departure from existing best-response algorithms, whose convergence conditions imply the uniqueness of the NE. Some of our results hinge on the use of VI problems directly in the complex domain; the study of these new kind of VIs also represents a noteworthy innovative contribution. We then apply the developed methods to solve some new generalizations of SISO and MIMO games in cognitive radios and femtocell systems, showing a considerable performance improvement over classical pure noncooperative schemes.

1 Introduction and Motivation

In recent years, there has been a growing interest in the use of noncooperative games to model and solve resource allocation problems in communications and networking, wherein the interaction among several agents is by no means negligible and centralized approaches are not suitable. Examples are power control and resource sharing in wireless/wired peer-to-peer networks, cognitive radio systems (e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]), distributed routing, flow and congestion control, and load balancing in communication networks (e.g., [13, 14, 15] and references therein), and smart grids (see [16, 17] and references therein). Two recent special issues on the subject are [18, 19].

*The work of J.-S. Pang is based on research supported by the U.S. National Science Foundation grant No. CMMI 0969600. The work of G. Scutari is supported by the National Science Foundation Grant No. CNS-1218717. The work of D. P. Palomar was supported by the Hong Kong RGC 617810 research grant. Part of this paper has been presented at the 31st Annual IEEE International Conference on Computer Communications (IEEE INFOCOM 2012).

Among the variety of models and solution concepts proposed in the literature, the Nash Equilibrium Problem (NEP) plays a central role and has been used most to model interactions among individuals competing selfishly for scarce resources. In a NEP there is a finite number I of players; each player i makes decisions on a set of variables \mathbf{x}_i belonging to a given feasible set $\mathbf{x}_i \in \mathcal{Q}_i$. The goal of each player i is to minimize his own objective function $f_i(\mathbf{x}_i, \mathbf{x}_{-i})$ over \mathcal{Q}_i while anticipating the reactions $\mathbf{x}_{-i} \triangleq (\mathbf{x}_j)_{j \neq i=1}^I$ from the rivals:

$$\begin{aligned} & \underset{\mathbf{x}_i}{\text{minimize}} && f_i(\mathbf{x}_i, \mathbf{x}_{-i}) \\ & \text{subject to} && \mathbf{x}_i \in \mathcal{Q}_i. \end{aligned} \tag{1}$$

The NEP is the problem of finding a vector $\mathbf{x}^* \triangleq (\mathbf{x}_i^*)_{i=1}^I$ such that each \mathbf{x}_i^* belongs to \mathcal{Q}_i and solves the player's problem (given \mathbf{x}_{-i}^*):

$$f_i(\mathbf{x}_i^*, \mathbf{x}_{-i}^*) \leq f_i(\mathbf{x}_i, \mathbf{x}_{-i}^*), \quad \forall \mathbf{x}_i \in \mathcal{Q}_i. \tag{2}$$

Such a point \mathbf{x}^* is called a Nash Equilibrium (NE) or, more simply, a solution of the NEP. In words, a NE is a feasible strategy profile \mathbf{x}^* such that no *single* player can benefit from a *unilateral* deviation from \mathbf{x}_i^* .

In this paper we focus on NEPs in the general form (1), in the following setting: i) the optimization variables of each player can be either real vectors or complex matrices; ii) each optimization problem in (1) is convex for any given feasible \mathbf{x}_{-i} ; and iii) players' objective functions are continuously differentiable in all the variables (more precisely, functions of complex variables are assumed to be \mathbb{R} -differentiable, see Sec. 6). We will term such a game (*real* or *complex*) *player-convex* NEP. Note that assumptions ii) and iii) are mild and quite standard in the literature, see for example [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 18, 19] where special instances of the player-convex NEP (1) are studied. The convexity assumption ii) makes the NEP numerically tractable (a NE may not even exist otherwise) while, to date, the differentiability of the players' functions seems indispensable to analyze *distributed* solution methods [20, 21] (unless the game has a very specific structure, like in potential or supermodular games; see, e.g., [22, 23] and references therein). Motivated by recent applications of noncooperative models in MIMO communications [7, 8, 10, 11, 24], we also allow, according to i), players' optimization variables to be complex matrices, which significantly enlarges the range of applicability of model (1). To the best of our knowledge, this is the first work where a *complex* NEP in the general form (1) is considered.

While the solution analysis (existence, etc.) of a *real* player-convex NEP relies on standard results in game theory (see, e.g., the seminal work [25], or [20] for more recent results), the development of *distributed* solution algorithms is much more involved. The goal of this paper is to address this difficult task in the broad setting described above. We are interested in the design and analysis of (possibly) asynchronous iterative *best-response* algorithms, suitable for solving real and complex player-convex NEPs, even in the presence of multiple NEs. By "best-response" algorithms we mean iterative schemes where the players iteratively choose the (feasible) strategy that minimizes their cost functions, given the actions of the other players; the reason for our emphasis on best-response schemes will be described shortly.

1.1 Literature review

The study of iterative algorithms for (special cases of) player-convex NEPs has been addressed in a number of papers, under different settings and assumptions; the main features and limitations of current state-of-the-art approaches are discussed next.

A first class of papers is composed of works motivated by specific applications, some examples are [1, 2, 3, 4, 5, 7, 8, 9, 10, 11], where different resource allocation problems in communications are modelled as noncooperative games and solved via iterative algorithms; all these formulations are special cases of the NEP (1). A key feature of all these models is that the best-response of each player (i.e., the optimal solution of each player's optimization problem) is unique and can be expressed in closed form; this simplifies enormously the application of standard fixed-point arguments to the study of the convergence of best-response algorithms. The same line of analysis is developed also in [26] and [27] for quadratic and non-quadratic (but with only two-players) real convex NEPs, respectively. A monotonicity-based approach is instead used in [28, 29]. Even though algorithms in [26, 27, 28] do not require a closed form solution of players' optimization problems, they can be computationally very demanding and the convergence conditions are based on assumption whose verification for games arising from realistic applications remains elusive. Last but not least, convergence conditions of the algorithms proposed in all the aforementioned papers imply the uniqueness of the NE.

A more general and powerful methodology suitable for studying noncooperative games is offered by the theory of finite-dimensional Variational Inequalities (VIs) [30]. VI and complementarity problems have a long history and have been well documented in the literature of operation research [30], but only recently they have been brought also to the attention of the signal processing, communications, and networking communities [2, 4, 6, 10, 31, 32]. Given a subset \mathcal{K} of \mathbb{R}^n and a vector-valued function $\mathbf{F} : \mathcal{K} \rightarrow \mathbb{R}^n$, the VI problem, denoted by $\text{VI}(\mathcal{K}, \mathbf{F})$, consists in finding a point $\mathbf{x}^* \in \mathcal{K}$ such that

$$(\mathbf{x} - \mathbf{x}^*)^T \mathbf{F}(\mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{K}. \quad (3)$$

The VI approach to noncooperative games hinges on an easy equivalence between *real* player-convex NEPs as in (1) and the (partitioned) VI problem $\text{VI}(\mathcal{K}, \mathbf{F})$ in (3), with $\mathcal{K} = \prod_{i=1}^I \mathcal{Q}_i$ and $\mathbf{F} = (\nabla_{\mathbf{x}_i} f_i(\mathbf{x}))_{i=1}^I$ (intended to be a column vector), where $\nabla_{\mathbf{x}_i} f_i(\mathbf{x})$ denotes the gradient of f_i with respect to \mathbf{x}_i . Based on this equivalence, one can solve a real player-convex NEP by focusing on the associated VI problem and taking advantage of the many (centralized and distributed) solution methods available in the literature for partitioned VIs [30, Vol. II].

In the effort of obtaining distributed schemes for NEPs, researchers have focused on so called *projection algorithms* for partitioned VIs [30, Ch. 12]. When applied to NEPs, projection algorithms work as follows: at each iteration, all the players choose simultaneously their own strategies employing a (projected) gradient update. This approach has been followed in a number of works, e.g., [33, 34, 35]; continuous-time strategy evolutions based on players' gradient responses have been studied also in [25, 36] (without using the VI formalism). Even though algorithms in [33, 34, 35] seem to be a most natural choice to solve NEPs while using existing VI solution methods, they suffer, together with [25, 36], from some drawbacks, which strongly limit their applicability in practice. First, they are not "incentive compatible", meaning that selfish users may deviate from them, unless they are imposed by some authority as a protocol to follow. Second, and most importantly, they generally converge very slowly; this has been observed in a number of different applications (see, e.g., numerical results in [34, 35, 36]). For instance, our experiments show that gradient-response algorithms need about two orders of magnitude more iterations than best-response schemes to reach comparable performance (see, Figure 6 in Sec. 8); this makes them not appealing in practice, especially for the design of wireless systems, where iterations correspond to physical transmissions.

A different approach to the design of algorithms for partitioned VIs has been followed in [37, 38], where the authors investigated the local and global convergence of various iterative synchronous methods that decompose the original VI problem into a sequence of simpler lower-dimensional VI subproblems. When the VI problem comes from a real player-convex NEP, some of the iterative algorithms in [37, 38] can be interpreted as synchronous best-response schemes solving the NEP. Unfortunately the convergence analysis in [37, 38], based on contraction arguments, leads to abstract convergence conditions, whose verification in practice seems not possible. Easier conditions to be checked have been obtained recently in [20] for simultaneous best-response algorithms, still using the VI approach. However, conditions in [20, 37, 38] are applicable only to a restricted class of *real* NEPs; they indeed imply the (uniformly) strongly convexity of the players' cost functions and the uniqueness of the NE. In the presence of multiple solutions, the distributed computation of even a single NE of *real/complex* NEPs via best-response algorithms becomes a difficult and unsolved task.

The analysis of the current literature carried out so far leads to the following conclusions: When it comes to distributed computation of NE via best-response dynamics, the following issues arise: i) the convergence analysis and algorithms apply only to a restricted class of NEPs, whose players' cost functions and feasible sets have a very specific structure, leaving outside player-convex NEPs in the general form (1); ii) the best-response mapping of each player must be unique and/or is required to be computed in closed form; iii) convergence is obtained only under conditions implying the uniqueness of the NE; and iv) none of current results and VI-based methodologies can be applied to study and solve *complex* player-convex NEPs, which arise naturally, e.g., from applications in MIMO communications.

1.2 Main contributions

In order to address the key issues listed at the end of the previous subsection, in this paper we introduce several new developments that are summarized next.

1. Building on our recent contributions [20, 21, 31, 39], we develop a VI-based unified theory for the study and design of distributed best-response algorithms for the solution of *real* player-convex NEPs, having (possibly) multiple solutions. Our unified framework has many desirable properties, such as:
 - It provides a systematic methodology for analyzing old and new algorithms, simplifying greatly the application of game-theoretical models to new problems;
 - It includes (both synchronous and) totally asynchronous methods, thus improving on the more traditional synchronous methods studied in the literature, see e.g. [20, 25, 33, 34, 35, 36]). Note that in many applications asynchronous algorithms play a fundamental role and are the only methods that can reasonably be used;
 - It advances the state-of-the-art algorithms by relaxing the convergence conditions. In particular, the proposed distributed best-response schemes are suitable to solve player-convex NEPs having multiple solutions; moreover, if some (limited) signaling among the players is allowed, one can also control the quality of the achievable solution by forcing convergence to a NE that optimizes a further performance criterion (i.e., performing an equilibrium selection). This feature is very appealing in the design of practical wireless systems, where algorithms with unpredictable performance (like [1, 2, 3, 4, 5, 7, 8, 9,

10, 11] and [25, 33, 34, 35, 36], when multiple solutions are present) are not acceptable;

- It does not require the players’ best-response to be unique or given in closed form;
- It allows us to clearly gauge the trade-off between signaling and characteristic of the resulting algorithms.

2. We develop an entirely new theory for the study of VIs in the complex domain, which allows us to extend all the results we established in the real case to player-convex NEPs whose players’ optimization variables are complex matrices. Key to this extension is the development of several new technical results that are of independent interest and useful in many other applications. Among these we single out

- A Taylor expansion of real-valued functions of complex matrices that is amenable to our MIMO applications;
- The minimum principle for constrained convex optimization problems in the domain of complex matrices, generalizing the already known complex gradient-vanishing conditions obtained in [40] for the *unconstrained* minimization of real-valued functions of complex variables; and
- A new definition of Hessian matrix for real-valued functions of complex matrices (and Jacobian matrix of complex-valued VI functions) along with a relaxed concept of positive (semi-)definiteness, which are instrumental to establish the connection with the convexity properties of the function (and the monotonicity/P properties of the complex-valued VI function).

To the best of our knowledge the above features constitute a substantial advancement in the distributed solution methods of noncooperative games, which enlarges considerably scope and flexibility of game-theoretical models in wireless distributed (MIMO) networks. We believe that the results in this paper bring the use of game-theoretic methods one step closer to “technological” maturity and will facilitate further developments and new applications in several fields.

In order to illustrate our techniques we consider some new MIMO games over vector Gaussian Interference Channels (ICs), modeling some distributed resource allocation problems in SISO and MIMO CR systems and femtocells. These games are examples of NEPs that cannot be handled by current methodologies. The application of our framework leads to a unified algorithm converging to different types of solutions according to the selected degree of cooperation and signaling (in the form of pricing at the level of players’ objective functions). The choice of the specific (pricing) scheme will depend then on the trade-off between signaling and performance that the users are willing to exchange/achieve. Numerical results show the superiority of our approach with respect to plain noncooperative solutions as well as good performance with respect to centralized solutions, in favor of very limited signaling among the players.

The paper is organized as follows. Sec. 2 introduces the just mentioned new resource allocations problems. Sec. 3 presents some results on partitioned VIs geared towards the analysis of NEPs; special emphasis is given to some classes of vector functions \mathbf{F} and its properties that play a key role in the solution analysis of the VI as well as convergence analysis of distributed algorithms for NEPs. Establishing the connection between partitioned VIs and real convex-player NEPs and building on results introduced in Sec. 3, Sec. 4 focuses on the solution analysis of real convex-player NEPs (more general results are given in Appendix A). Sec. 5 and Sec. 6 constitute the core theoretical part of the paper; in Sec. 5 we provide various distributed

algorithms for solving real player-convex NEPs in several significant settings along with their convergence properties; Sec. 6 generalizes the main results in Sec. 3, 4, and 5 on real VIs to the complex case. Sec. 7 shows how to apply the developed machinery to the concrete resource allocation problems introduced in Sec. 2, whereas Sec. 8 provides some numerical results corroborating our theoretical findings. Finally, Sec. 9 draws some conclusions. The proof of the major results of the paper is given in Appendix A-I.

2 Motivating Examples: Noncooperative Games Over Gaussian ICs

To motivate and illustrate our new results more in detail, we start introducing some novel resource allocation problems over SISO frequency-selective and MIMO Gaussian ICs, widely extending formulations that have already been studied in the literature. We will show that these problems cannot be analyzed and solved using current results and algorithms, but call for a more general theory.

The IC is suitable to model many practical multiuser systems, such as digital subscriber lines, wireless ad-hoc and Cognitive Radio (CR) networks, peer-to-peer systems, multicell OFDM/TDMA cellular systems, and Femtocell-based networks. To have specific case studies at hand, we will focus on CR systems and Femtocell-based networks; the proposed techniques can be readily applied also to the other aforementioned network models.

2.1 The SISO case

We consider an I -user N -parallel Gaussian interference channel, modeling a CR system composed of Q secondary users (SUs) and P primary users (PUs). In this model, there are I transmitter-receiver pairs—the SUs—where each transmitter wants to communicate with its corresponding receiver over a set of N parallel Gaussian subchannels which may represent time or frequency bins (here we consider transmissions over the frequency-selective IC without loss of generality). We denote by $H_{ij}(k)$ the (cross-) channel transfer function over the k -th frequency bin between the secondary transmitter j and the receiver i , while the channel transfer function of secondary link i is $H_{ii}(k)$. The transmission strategy of each user (pair) i is the power allocation vector $\mathbf{p}_i = \{p_i(k)\}_{k=1}^N$ over the N subcarriers; the power budget of each transmitter i is $\sum_{k=1}^N p_i(k) \leq P_i$. In a CR system, additional power constraints limiting the interference radiated by the SUs need to be imposed. Here we envisage the use of the following general interference constraints: for each SU i ,

$$\sum_{k=1}^N \mathbf{w}_i(k) p_i(k) \leq \boldsymbol{\alpha}_i, \quad i = 1, \dots, I, \quad (4)$$

where $\mathbf{w}_i(k) \in \mathbb{R}_+^m$ and $\boldsymbol{\alpha}_i \in \mathbb{R}_+^m$ are nonnegative m -length vectors. Note that constraints in the form of (4) are general enough to include, as special cases, for example: i) spectral mask constraints $0 \leq \mathbf{p}_i \leq \mathbf{p}_i^{\max}$, where $\mathbf{p}_i^{\max} = (p_i^{\max}(k))_{k=1}^N$ is the vector of spectral masks over licensed bands; and ii) interference temperature limit-like constraints $\sum_{k=1}^N |H_{pi}^{(P,S)}(k)|^2 p_i(k) \leq I_{pi}$ for $p = 1, \dots, P$, where $H_{pi}^{(P,S)}(k)$ is the cross-channel transfer function over carrier k between the secondary transmitter i and the primary receiver p , and I_{pi} is the maximum level of interference that SU i is allowed to generate. Methods to obtain the interference thresholds I_{pi} when such a knowledge is not available at the SUs' side are discussed in [41, Sec. IV]. For the purposes

of this section, define by

$$\tilde{\mathcal{P}}_i^{\text{siso}} \triangleq \left\{ \mathbf{p}_i \in \mathbb{R}^N : \sum_{k=1}^N p_i(k) \leq P_i, \quad \mathbf{0} \leq \mathbf{p}_i \leq \mathbf{p}_i^{\max} \right\}, \quad (5)$$

the set of power budget constraint of SU i including explicitly the power budget and spectral mask constraints.

Under basic information theoretical assumptions (see, e.g., [1, 4]), the maximum achievable rate on link i for a specific power allocation profile $\mathbf{p}_1, \dots, \mathbf{p}_I$ is

$$r_i(\mathbf{p}_i, \mathbf{p}_{-i}) = \sum_{k=1}^N \log \left(1 + \frac{|H_{ii}(k)|^2 p_i(k)}{\sigma_i^2(k) + \sum_{j \neq i} |H_{ij}(k)|^2 p_j(k)} \right) \quad (6)$$

where $\mathbf{p}_{-i} \triangleq (\mathbf{p}_1, \dots, \mathbf{p}_{i-1}, \mathbf{p}_{i+1}, \dots, \mathbf{p}_I)$ is the set of all the users power allocation vectors, except the i -th one, and $\sigma_i^2(k) + \sum_{j \neq i} |H_{ij}(k)|^2 p_j(k)$ is the variance of the noise plus the multiuser interference (MUI) over subcarrier k measured by the receiver i , with $\sigma_i^2(k)$ denoting the power of the thermal noise (possibly including the interference generated by the PUs). The system design consists in finding the optimal power allocation of the users in order to maximize the information rates of the links, according to some performance metrics. The sum-rate maximization [in the absence of the interference constraints (4)] has been shown to be an NP hard problem [42]; several attempts have been pursued in the literature to deal with the nonconvexity of such a problem; however all the proposed schemes are centralized and computationally expensive, which makes them non-implementable in a network with no infrastructure, like ad-hoc or CR networks. This motivates the formulation of the system design as a NEP: the aim of each player (link) i , given the strategy profile \mathbf{p}_{-i} of the others, is to choose a feasible power allocation \mathbf{p}_i that maximizes the rate $r_i(\mathbf{p}_i, \mathbf{p}_{-i})$, i.e.,

$$\begin{aligned} & \underset{\mathbf{p}_i}{\text{maximize}} && r_i(\mathbf{p}_i, \mathbf{p}_{-i}) \\ & \text{subject to} && \\ & \quad \text{(a)} : && \mathbf{p}_i \in \tilde{\mathcal{P}}_i^{\text{siso}}, \\ & \quad \text{(b)} : && \sum_{k=1}^N \mathbf{w}_i(k) p_i(k) \leq \alpha_i, \end{aligned} \quad \left. \vphantom{\sum_{k=1}^N} \right\} \triangleq \mathcal{P}_i^{\text{siso}} \quad (7)$$

for all $i = 1, \dots, I$, where $\tilde{\mathcal{P}}_i^{\text{siso}}$ and $r_i(\mathbf{p}_i, \mathbf{p}_{-i})$ are defined in (5) and (6), respectively. We denote the NEP based on (7) by $\mathcal{G}_{\text{siso}} = \langle \mathcal{P}^{\text{siso}}, (r_i)_{i=1}^I \rangle$, with $\mathcal{P}^{\text{siso}} \triangleq \prod_i \mathcal{P}_i^{\text{siso}}$ and $\mathcal{P}_i^{\text{siso}}$ being the feasible set of the optimization problem (7) of SU i . Note that $\mathcal{G}_{\text{siso}}$ is an instance of the real player-convex NEP in (1).

Literature review. Special cases of the NEP in (7) have been extensively studied in the literature in the context of ad-hoc networks, namely when there are *only power constraints* (a) [1, 2, 4, 5, 43]. In such a simplified setting, given the strategy profile \mathbf{p}_{-i} , the optimization problem of each player reduces to:

$$\begin{aligned} & \underset{\mathbf{p}_i}{\text{maximize}} && r_i(\mathbf{p}_i, \mathbf{p}_{-i}) \\ & \text{subject to} && \mathbf{p}_i \in \tilde{\mathcal{P}}_i^{\text{siso}}. \end{aligned} \quad (8)$$

We denote the game resulting from (8) by $\tilde{\mathcal{G}}_{\text{siso}} = \langle \tilde{\mathcal{P}}^{\text{siso}}, (r_i)_{i=1}^I \rangle$, with $\tilde{\mathcal{P}}^{\text{siso}} \triangleq \prod_i \tilde{\mathcal{P}}_i^{\text{siso}}$. Introducing the matrices $\mathbf{M} \triangleq (\mathbf{M}_{ij})_{i,j=1}^I \in \mathbb{R}^{N \times I \times N \times I}$ and $\mathbf{\Gamma} \in \mathbb{R}^{I \times I}$ defined respectively as

$$\mathbf{M}_{ij} \triangleq \text{diag} \left\{ \left(\frac{|H_{ij}(k)|^2}{|H_{ii}(k)|^2} \right)_{k=1}^N \right\} \quad \text{and} \quad [\mathbf{\Gamma}]_{ij} \triangleq \begin{cases} 0, & \text{if } i = j; \\ \max_k \frac{|H_{ij}(k)|^2}{|H_{ii}(k)|^2}, & \text{otherwise,} \end{cases} \quad (9)$$

the state-of-the-art-results on $\tilde{\mathcal{G}}_{\text{siso}}$ can be collected together in the following theorem, where $\rho(\mathbf{A})$ denotes the spectral radius of \mathbf{A} .

Theorem 1 *Given the NEP $\tilde{\mathcal{G}}_{\text{siso}}$ (with no interference constraints), the following hold.*

- (a) $\tilde{\mathcal{G}}_{\text{siso}}$ has a nonempty and compact solution set;
- (b) If \mathbf{M} is positive definite, then $\tilde{\mathcal{G}}_{\text{siso}}$ has a unique NE [2, 4];
- (c) If $\rho(\mathbf{\Gamma}) < 1$, then $\tilde{\mathcal{G}}_{\text{siso}}$ has a unique NE and the Asynchronous Iterative Waterfilling Algorithm (IWFA) based on the waterfilling best-response as proposed in [4] converges to the equilibrium.

Theorem 1 provides a satisfactory characterization of the NEP $\tilde{\mathcal{G}}_{\text{siso}}$ (namely, conditions for the existence/uniqueness of the solution and global convergence of distributed solution schemes) under $\rho(\mathbf{\Gamma}) < 1$ (or \mathbf{M} positive definite). However, condition $\rho(\mathbf{\Gamma}) < 1$ may be too restrictive in practice; indeed there are channel scenarios resulting in games $\tilde{\mathcal{G}}_{\text{siso}}$ having multiple Nash equilibria for which $\rho(\mathbf{\Gamma}) < 1$ does not hold. In such cases, the IWFA is no longer guaranteed to converge and there are no algorithms available in the literature solving the game $\tilde{\mathcal{G}}_{\text{siso}}$. Moreover, results in Theorem 1 as well as the mathematical tools used in [1, 2, 4, 5, 43] to study $\tilde{\mathcal{G}}_{\text{siso}}$ cannot be applied to the more general $\mathcal{G}_{\text{siso}}$, even in the case of unique NE. The theoretical analysis of $\mathcal{G}_{\text{siso}}$ is then an open problem, which will be addressed in Sec. 7, based on the general framework that we introduce in the forthcoming sections. It is worth mentioning that we also derive a closed form waterfilling-like expression of the best-response in (7), which is a new results in the literature of communications.

2.2 The MIMO case

In a MIMO setting, the secondary terminals are equipped with multiple transmit/receive antennas and are allowed to transmit over a multidimensional space, whose coordinates may represent time slots, frequency bins, and angles. The goal is to find out the most appropriate transmission strategy exploring all available degrees of freedom, under alternative interference constraints (in addition to arbitrary power constraints imposed on the transmit covariance matrix of the SUs). We envisage the use of interference constraints expressed in the following very general form:

- *Null shaping constraints:*

$$\mathbf{U}_i^H \mathbf{Q}_i = \mathbf{0},$$

where $\mathbf{Q}_i \in \mathbb{C}^{n_{T_i} \times n_{T_i}}$ is the transmit covariance matrix of SU i with n_{T_i} being the number of transmit antennas and $\mathbf{U}_i \in \mathbb{C}^{n_{T_i} \times r_{U_i}}$ is a tall matrix whose columns represent the “directions” along with user q is not allowed to transmit. We assume, without loss of generality (w.l.o.g.) that each matrix \mathbf{U}_i is full-column rank and, to avoid the trivial solution $\mathbf{Q}_i = \mathbf{0}$, $r_{U_i} < n_{T_i}$.

- *Soft and peak power shaping constraints:*

$$\text{tr}(\mathbf{G}_{pi}^H \mathbf{Q}_i \mathbf{G}_{pi}) \leq I_{pi}^{\text{ave}} \quad \text{and} \quad \lambda_{\max}(\mathbf{F}_{pi}^H \mathbf{Q}_i \mathbf{F}_{pi}) \leq I_{pi}^{\text{peak}}, \quad p = 1, 2, \dots,$$

which represent a relaxed version of the null constraints by limiting the total average and peak average power radiated along the range space of matrix $\mathbf{G}_{pi} \in \mathbb{C}^{n_{T_i} \times n_{G_p}}$ and $\mathbf{F}_{pi} \in \mathbb{C}^{n_{T_i} \times n_{F_p}}$, where I_{pi}^{ave} and I_{pi}^{peak} are the maximum average and average peak power respectively that can be transmitted along the directions spanned by \mathbf{G}_{pi} and \mathbf{F}_{pi} .

Null constraints are enforced to prevent SUs from transmitting over prescribed subspaces, which for example can identify: 1) portion of licensed spectrum (the range space of \mathbf{U}_i coincides with the subspace spanned by a set of Inverse Discrete Fourier Transform (IDFT) vectors); 2) time slots used by the PUs (the columns of \mathbf{U}_i are a set of canonical vectors); and 3) angular directions identifying the primary receivers as observed from the secondary transmitters (the columns of \mathbf{U}_i are the set of steering vectors representing the directions of the primary receivers as observed from the secondary transmitters). Soft shaping constraints can be used instead to control the (average and peak average) power radiated by the SUs along prescribed time/frequency/angular “directions” (those spanned by the columns of matrices \mathbf{G}_{pi} and \mathbf{F}_{pi}); for instance, classical power constraints, such as per-antenna power constraints $[\mathbf{Q}_i]_{kk} \leq \beta_{ik}$ with $k = 1, \dots, n_{T_i}$, or power budget constraints $\text{tr}(\mathbf{Q}_i) \leq P_i$ are example of soft-shaping constraints. In the CR context, null constraints are motivated by the interference-avoiding paradigm (also called white-space filling approach): CR nodes sense the spatial, temporal, or spectral voids and adjust their transmission strategy to fill in the sensed white spaces; such a paradigm has already been adopted as a core platform in emerging wireless access standards such as the IEEE 802.22-Wireless Regional Area Networks (WRANs). Soft-shaping constraints find their natural application within the opportunistic communication paradigm, which allows simultaneous transmissions between primary and secondary users, provided that the required QoS of the PUs is preserved [41]. Under basic information theoretical assumptions (see, e.g., [8]), the maximum information rate on secondary link i for a given set of user covariance matrices $\mathbf{Q}_1, \dots, \mathbf{Q}_I$, is

$$R_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) = \log \det (\mathbf{I} + \mathbf{H}_{ii}^H \mathbf{R}_{-i}(\mathbf{Q}_{-i})^{-1} \mathbf{H}_{ii} \mathbf{Q}_i) \quad (10)$$

where $\mathbf{R}_{-i}(\mathbf{Q}_{-i}) \triangleq \mathbf{R}_{n_i} + \sum_{j \neq i} \mathbf{H}_{ij} \mathbf{Q}_j \mathbf{H}_{ij}^H$ is the covariance matrix of the noise plus MUI, with $\mathbf{R}_{n_i} \in \mathbb{C}^{n_{R_i} \times n_{R_i}}$ denoting the covariance matrix of the thermal Gaussian zero mean noise (possibly including the interference generated by the PUs), and assumed to be positive definite; $\mathbf{Q}_{-i} \triangleq (\mathbf{Q}_j)_{j \neq i}$ is the set of all the users covariance matrices, except the i -th one; $\mathbf{H}_{ii} \in \mathbb{C}^{n_{R_i} \times n_{T_i}}$ is the channel matrix between the i -th secondary transmitter and the intended receiver, whereas $\mathbf{H}_{ij} \in \mathbb{C}^{n_{R_i} \times n_{T_j}}$ is the cross-channel matrix between secondary source j and destination i . Within the above setup, the game theoretical formulation is: for each SU $i = 1, \dots, I$,

$$\begin{aligned} & \underset{\mathbf{Q}_i \succeq \mathbf{0}}{\text{maximize}} && R_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) \\ & \text{subject to} && \\ & \quad \text{(a) :} && \text{tr}(\mathbf{Q}_i) \leq P_i, \\ & \quad \text{(b) :} && \mathbf{U}_i^H \mathbf{Q}_i = \mathbf{0}, \\ & \quad \text{(c) :} && \text{tr}(\mathbf{G}_{pi}^H \mathbf{Q}_i \mathbf{G}_{pi}) \leq I_{pi}^{\text{ave}}, \quad \lambda_{\max}(\mathbf{F}_{pi}^H \mathbf{Q}_i \mathbf{F}_{pi}) \leq I_{pi}^{\text{peak}}, \quad p = 1, 2, \dots, \\ & \quad \text{(d) :} && \mathbf{Q}_i \in \mathcal{Q}_i, \end{aligned} \quad \left. \vphantom{\begin{aligned} & \underset{\mathbf{Q}_i \succeq \mathbf{0}}{\text{maximize}} \\ & \text{subject to} \\ & \quad \text{(a) :} \\ & \quad \text{(b) :} \\ & \quad \text{(c) :} \\ & \quad \text{(d) :} \end{aligned}} \right\} \triangleq \mathcal{P}_i^{\text{mimo}} \quad (11)$$

where $\mathcal{Q}_i \subseteq \mathbb{C}^{n_{T_i} \times n_{T_i}}$ is an abstract set that can accommodate (possibly) additional constraints on the covariance matrix \mathbf{Q}_i , on top of the power and interference constraints; we only make the (blanket) assumption that each \mathcal{Q}_i is closed and convex. We refer to the NEP based on (11) as $\mathcal{G}_{\text{mimo}} = \langle \mathcal{P}^{\text{mimo}}, (R_i)_{i=1}^I \rangle$, with $\mathcal{P}^{\text{mimo}} \triangleq \prod_i \mathcal{P}_i^{\text{mimo}}$ and $\mathcal{P}_i^{\text{mimo}}$ defined in (11). Note that $\mathcal{G}_{\text{mimo}}$ is an instance of the complex NEP (1).

Literature review. The design of MIMO CR systems under different interference-power/interference-temperature constraints has been addressed in a number of papers. An overview of centralized solutions can be found in [44], whose approach is to cast the resource allocation of the secondary system into a classical *convex* optimization problem, which is suitable for modeling CR networks composed of *one* SU only (the case of multiple SUs would lead to a NP-hard formulation). Distributed algorithms (mostly) for ad-hoc networks based on game theoretical formulations have been proposed instead in [24, 7, 45, 8, 11]; the state-of-the-art result is the asynchronous MIMO IWFA solving the NEP in (11), in the presence of constraints (a) [8] and (b) [11] only. Results in these papers are strongly based on the specific structure of the optimization problem and the resulting solution—the MIMO waterfilling-like expression—and thus are not applicable to the general NEP (11). $\mathcal{G}_{\text{mimo}}$ is thus an other example of a novel game whose solution analysis requires new mathematical tools, which is the goal of this paper. The study of $\mathcal{G}_{\text{mimo}}$ is addressed in Sec. 7.2 and will result as a direct application of the framework developed in the forthcoming sections for complex NEPs.

2.3 Femtocell OFDM MIMO networks

Femtocell networks [46] are two-tier systems composed of central macrocell base stations [also termed evolved Node-Bs (eNBs) in the 3GPP jargon] and multiple distributed femtocells [also termed as Home evolved Node-Bs (HeNBs)]; the latter being short-range, low-power base stations that are deployed and managed by the customers at home or in their offices. Due to the massive and dense deployment of HeNBs and their uncoordinated nature, interference management is one of the major challenges to be faced in the femtocell system design; different interference models and resource allocation algorithms have been proposed in the literature (see, e.g., [47, 48, 49] and references therein). Here we build on the interference model proposed in [49] and introduce a novel spectrum sharing resource allocation and interference management problem in OFDM downlink MIMO LTE femtocell systems. Borrowing the idea of CR systems described in the previous section, we formulate the design of the radio interference-aware access scheme for HeNBs as a complex player-convex NEP, where the eNBs play the role of interference (the PUs) and the HeNBs, acting as SUs, compete to optimize their available time/frequency/space degrees of freedom, based on a preliminary sensing and prediction of the interference activities of the eNBs.

Assuming the HeNBs to operate over a frame-base structure (the timeline is divided into T successive non-overlapping time-slots), the interference activities of the eNBs over the frequency channels $k = 1, \dots, N$ are modeled as a set of statistically independent first-order homogeneous Discrete Time Markov Chains [49] (higher order Markov chain can also be considered without any further conceptual complication); let $\pi_I^{(k)}(t)$ and $\pi_B^{(k)}(t)$ denote the probability that the subcarrier k in the time-slot $t = 1, \dots, T$ is idle and busy, respectively; introducing the matrix $\mathbf{P}^{(k)}$ of transition probabilities over channel k , the time evolution of the state probabilities $\boldsymbol{\pi}^{(k)}(t) \triangleq [\pi_I^{(k)}(t), \pi_B^{(k)}(t)]^T$, given the channel state $\boldsymbol{\pi}^{(k)}(0)$ in the initial time-slot $t = 0$, is $\boldsymbol{\pi}^{(k)}(t) = \mathbf{P}^{(k)} \boldsymbol{\pi}^{(k)}(t-1)$. The transition probabilities in the matrix $\mathbf{P}^{(k)}$ can be estimate from the

HeNBs either through a knowledge of the traffic statistics of the eNBs, or by sensing the channels and using well-known estimation techniques (see, e.g., [50]), or by a combination of these. Hereafter, we assume that the HeNBs allocate some of the available slots, let us say the first t_0 time slots of the transmission frame, to measure the received interference generated by the active HeNBs and eNBs over each carrier and estimate the transition probability matrices $\mathbf{P}^{(k)}$.

Based on such an estimate and the prediction model $\boldsymbol{\pi}^{(k)}(t) = \mathbf{P}^{(k)}\boldsymbol{\pi}^{(k)}(t-1)$ of the interference activities, the maximum achievable rate on the (downlink) MIMO OFDM channel of HeNB i over $T - t_0$ consecutive time-slots is:

$$\bar{R}_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) = \sum_{t=t_0+1}^T \sum_{k=1}^N \left(\pi_I^{(k)}(t) \bar{r}_{i,I}^{(k)}(\mathbf{Q}_i^{(k,t)}, \mathbf{Q}_{-i}^{(k,t)}) + \pi_B^{(k)}(t) \bar{r}_{i,B}^{(k)}(\mathbf{Q}_i^{(k,t)}, \mathbf{Q}_{-i}^{(k,t)}) \right) \quad (12)$$

with $\mathbf{Q}_i \triangleq ((\mathbf{Q}_i^{(k,t)})_{k=1}^N)_{t=t_0}^T$, and $\bar{r}_{i,I}^{(k)}(\mathbf{Q}_i^{(k,t)}, \mathbf{Q}_{-i}^{(k,t)})$ and $\bar{r}_{i,B}^{(k)}(\mathbf{Q}_i^{(k,t)}, \mathbf{Q}_{-i}^{(k,t)})$ being the achievable rates on HeNB link i over carrier k in the time slot t when the channel k is detected as idle and busy, respectively:

$$\begin{aligned} \bar{r}_{i,I}^{(k)}(\mathbf{Q}_i^{(k,t)}, \mathbf{Q}_{-i}^{(k,t)}) &\triangleq \log \det \left(\mathbf{I} + \mathbf{H}_{ii}^{(k)H} \mathbf{R}_{-i,I}^{(k)} (\mathbf{Q}_{-i}^{(k,t)})^{-1} \mathbf{H}_{ii}^{(k)} \mathbf{Q}_i^{(k,t)} \right) \\ \bar{r}_{i,B}^{(k)}(\mathbf{Q}_i^{(k,t)}, \mathbf{Q}_{-i}^{(k,t)}) &\triangleq \log \det \left(\mathbf{I} + \mathbf{H}_{ii}^{(k)H} \mathbf{R}_{-i,B}^{(k)} (\mathbf{Q}_{-i}^{(k,t)})^{-1} \mathbf{H}_{ii}^{(k)} \mathbf{Q}_i^{(k,t)} \right) \end{aligned}$$

where $\mathbf{R}_{-i,I}^{(k)}(\mathbf{Q}_{-i}^{(k,t)}) \triangleq \mathbf{R}_{n_i,I}^{(k)} + \sum_{j \neq i} \mathbf{H}_{ij}^{(k)} \mathbf{Q}_j^{(k,t)} \mathbf{H}_{ij}^{(k)H}$ and $\mathbf{R}_{-i,B}^{(k)}(\mathbf{Q}_{-i}^{(k,t)}) \triangleq \mathbf{R}_{n_i,B}^{(k)} + \sum_{j \neq i} \mathbf{H}_{ij}^{(k)} \mathbf{Q}_j^{(k,t)} \mathbf{H}_{ij}^{(k)H}$, with $\mathbf{R}_{n_i,I}^{(k)}$ and $\mathbf{R}_{n_i,B}^{(k)}$ denoting the covariance matrix of the thermal noise and the thermal noise plus the interference generated by the eNBs over carrier k at the receiver of the femto-user i , respectively; $\mathbf{Q}_i^{(k,t)}$ is the transmit covariance matrix of HeNB i over carrier k at time-slot t ; and $\mathbf{H}_{ij}^{(k)} \in \mathbb{C}^{n_{R_i} \times n_{T_j}}$ is the cross-channel matrix over carrier k between HeNB source j and the femto receiver i ($\mathbf{H}_{ii}^{(k)} \in \mathbb{C}^{n_{R_i} \times n_{T_i}}$ is the channel matrix over carrier k of the direct link i). All the channel matrices are assumed to be time-invariant over T consecutive time-slots. The goal of each HeNB i is then to maximize the expected MIMO femto-user rate (12), subject to power and interference constraints, the latter introduced to preserve the QoS of the users in the macro-cells. The proposed game-theoretical formulation is the following: for each HeNB i ,

$$\begin{aligned} &\underset{\mathbf{Q}_i \succeq \mathbf{0}}{\text{maximize}} \quad \bar{R}_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) \\ &\text{subject to} \quad \left. \begin{aligned} \text{(a)} : & \quad \sum_{t=t_0+1}^T \sum_{k=1}^N \text{tr}(\mathbf{Q}_i^{(k,t)}) \leq P_i, \quad \sum_{k=1}^N \text{tr}(\mathbf{Q}_i^{(k,t)}) \leq P_i^{(t)}, \quad \forall t > t_0, \\ \text{(b)} : & \quad \mathbf{U}_i^H \mathbf{Q}_i^{(k,t)} = \mathbf{0}, \quad \forall k \text{ and } t > t_0 \\ \text{(c)} : & \quad \text{tr}(\mathbf{G}_{pi}^{(k,t)H} \mathbf{Q}_i^{(k,t)} \mathbf{G}_{pi}^{(k,t)}) \leq I_{pi}^{\text{ave}}(k), \quad \lambda_{\max}(\mathbf{F}_{pi}^{(k,t)H} \mathbf{Q}_i^{(k,t)} \mathbf{F}_{pi}^{(k,t)}) \leq I_{pi}^{\text{peak}}(k), \quad \forall k, \forall t > t_0, \forall p \\ \text{(d)} : & \quad \sum_{k=1}^N \text{tr}(\mathbf{G}_{pi}^{(k,t)H} \mathbf{Q}_i^{(k,t)} \mathbf{G}_{pi}^{(k,t)}) \leq I_{pi}^{\text{ave}}, \quad \sum_{k=1}^N \lambda_{\max}(\mathbf{F}_{pi}^{(k,t)H} \mathbf{Q}_i^{(k,t)} \mathbf{F}_{pi}^{(k,t)}) \leq I_{pi}^{\text{peak}}, \quad \forall t > t_0, \forall p, \\ \text{(e)} : & \quad \mathbf{Q}_i \in \tilde{\mathcal{Q}}_i, \end{aligned} \right\} \triangleq \mathcal{P}_i^{\text{femto}} \quad (13) \end{aligned}$$

where $\tilde{\mathcal{Q}}_i$ is an abstract set of possibly additional constraints on $\mathbf{Q}_i = ((\mathbf{Q}_i^{(k,t)})_{k=1}^N)_{t=t_0}^T$, assumed to be closed and convex. The proposed formulation contains also per-carrier/whole-bandwidth average and peek average

power and interference constraints imposed to each HeNB i to protect the users in the macro-cells; the matrices $\mathbf{G}_{pi}^{(k,t)}$ and $\mathbf{F}_{pi}^{(k,t)}$ are chosen to select the “directions” along with the power radiated by each HeNB i needs to be limited, see Sec. 2.2 for a discussion and interpretation of these constraints. We refer to the NEP based on (13) as $\mathcal{G}_{\text{femto}} = \langle \mathcal{P}^{\text{femto}}, (\bar{R}_i)_{i=1}^Q \rangle$, with $\mathcal{P}^{\text{femto}} \triangleq \prod_i \mathcal{P}_i^{\text{femto}}$ and $\mathcal{P}_i^{\text{femto}}$ defined in (13). $\mathcal{G}_{\text{femto}}$ is another example of complex player-convex NEP in (1).

Literature review. The SISO formulation resulting as special case of (13) has been proposed in [49], in a simplified setting where power only budget constraints are considered. Based on [4], the authors in [49] derived sufficient conditions for the uniqueness of the NE and the convergence of Jacobi gradient-response algorithms. Convergence conditions in [49] are however not easy to be checked (they require an exhaustive search over the feasible set of the power constraints) and imply uniqueness of the equilibrium; moreover gradient-response schemes suffer from slow convergence (cf. Figure 6). The MIMO formulation (13) is instead novel and none of the approaches proposed in the literature of MIMO games [24, 7, 45, 8, 11] can be successfully applied to design best-response algorithms and study their convergence properties. The analysis of both SISO and MIMO formulations can be readily addressed using the framework that we introduce in this paper.

3 A Theory of (Partitioned) Variational Inequalities

Our study of NEPs is based on the reduction of a NEP to a VI. The main advantage of this reduction is algorithmic, since once the reduction has been carried out we can build on the well-developed VI theory [30] in order to design new solution methods for NEPs. A VI representing a NEP obviously has peculiarities that must be taken into account; the main purpose of this section then is to present some results on VIs geared towards the analysis of NEPs. Special emphasis will be given to VIs defined on a Cartesian product of sets, which we term *partitioned VIs*, as these will be widely used to study NEPs. We note that this section contains some results that are new even in the VI literature, such as new sufficient conditions for the block P property of a vector function.

The simplest way to see a VI is as a generalization of the minimum principle for convex optimization problems, which is recalled next. Consider a convex optimization problem (in the minimization form), whose objective function $f : \mathcal{Q} \mapsto \mathbb{R}$ is convex and continuously differentiable on the feasible set $\mathcal{Q} \subseteq \mathbb{R}^n$,¹ which is a convex and closed subset of \mathbb{R}^n . A feasible point $\mathbf{x}^* \in \mathcal{Q}$ is an optimal solution of the optimization problem if and only if

$$(\mathbf{x} - \mathbf{x}^*)^T \nabla f(\mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in \mathcal{Q}. \quad (14)$$

The VI problem is a generalization of the minimum principle (14) where the gradient ∇f is substituted by a general real-valued vector function \mathbf{F} . More formally, we have the following. Let $\mathcal{Q} \subseteq \mathbb{R}^n$ be a nonempty, closed, and convex set and let $\mathbf{F} : \mathcal{Q} \rightarrow \mathbb{R}^n$ be a vector-valued real function. The Variational Inequality VI $(\mathcal{Q}, \mathbf{F})$ is the problem of finding a feasible point $\mathbf{x}^* \in \mathcal{Q}$ such that [30, Def. 1.1.1]

$$(\mathbf{x} - \mathbf{x}^*)^T \mathbf{F}(\mathbf{x}^*) \geq 0, \quad \forall \mathbf{x} \in \mathcal{Q}. \quad (15)$$

¹When we say that a (vector-valued) function is continuous or continuously differentiable on a closed set we mean that the function is so on an open set containing the closed set.

The set of solutions to this problem is denoted by $\text{SOL}(\mathcal{Q}, \mathbf{F})$. As we mentioned earlier, a case that is relevant in the analysis of NEPs is that of partitioned VIs. This corresponds to the set \mathcal{Q} being a cartesian product of lower-dimensional sets: $\mathcal{Q} \triangleq \prod_{i=1}^I \mathcal{Q}_i$, with each $\mathcal{Q}_i \subseteq \mathbb{R}^{n_i}$ being nonempty, closed, and convex and with

$n \triangleq \sum_{i=1}^I n_i$. When this structure arises it will be quite natural to partition both \mathbf{F} and \mathbf{x} accordingly and therefore write $\mathbf{F}(\mathbf{x}) = (\mathbf{F}_i(\mathbf{x}))_{i=1}^I$ and $\mathbf{x} = (\mathbf{x}_i)_{i=1}^I$, where $\mathbf{F}_i : \mathcal{Q} \rightarrow \mathbb{R}^{n_i}$ is the i th-component block function of \mathbf{F} and $\mathbf{x}_i \in \mathbb{R}^{n_i}$ is the i th-component block of \mathbf{x} .

Several standard problems in nonlinear programming, game theory, and nonlinear analysis can be naturally formulated as a VI problem; many examples can be found in [30, Ch. 1], [32], and [33]. Below we summarize some known facts and definitions about VI and provide a few new results pertaining to the partitioned VI setting.

3.1 Solution analysis

In order to present results about the existence and the structure of the solution set of a VI, we introduce some function classes. Some of these classes will also play a crucial role in the algorithmic developments in later sections.

Definition 2 *A mapping $\mathbf{F} : \mathcal{Q} \rightarrow \mathbb{R}^n$, with \mathcal{Q} closed and convex, is said to be*

(i) *monotone on \mathcal{Q} if for all pairs \mathbf{x} and \mathbf{y} in \mathcal{Q} ,*

$$(\mathbf{x} - \mathbf{y})^T (\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})) \geq 0; \quad (16)$$

(ii) *strictly monotone if for all pairs $\mathbf{x} \neq \mathbf{y}$ in \mathcal{Q} the inequality in (16) is strict;*

(iii) *strongly monotone on \mathcal{Q} if there exists a constant $c_{\text{sm}} > 0$ such that for all pairs \mathbf{x} and \mathbf{y} in \mathcal{Q} ,*

$$(\mathbf{x} - \mathbf{y})^T (\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})) \geq c_{\text{sm}} \|\mathbf{x} - \mathbf{y}\|^2. \quad (17)$$

The constant c_{sm} is called strong monotonicity constant.

If we assume a Cartesian product structure, i.e. $\mathbf{F} = (\mathbf{F}_i(\mathbf{x}))_{i=1}^I$ and $\mathcal{Q} = \prod_i \mathcal{Q}_i$, then the function \mathbf{F} is said to be

(iv) *a P_0 function on \mathcal{Q} if for all pairs of distinct tuples $\mathbf{x} = (\mathbf{x}_i)_{i=1}^I$ and $\mathbf{y} = (\mathbf{y}_i)_{i=1}^I$ in \mathcal{Q} , an index i exists such that $\mathbf{x}_i \neq \mathbf{y}_i$ and*

$$(\mathbf{x}_i - \mathbf{y}_i)^T (\mathbf{F}_i(\mathbf{x}) - \mathbf{F}_i(\mathbf{y})) \geq 0; \quad (18)$$

(v) *a P function on \mathcal{Q} if for all pairs of distinct tuples $\mathbf{x} = (\mathbf{x}_i)_{i=1}^I$ and $\mathbf{y} = (\mathbf{y}_i)_{i=1}^I$ in \mathcal{Q} , the inequality in (18) is strict;*

(vi) a uniformly P function on \mathcal{Q} if there exists a constant $c_{\text{uP}} > 0$ such that for all pairs $\mathbf{x} = (\mathbf{x}_i)_{i=1}^I$ and $\mathbf{y} = (\mathbf{y}_i)_{i=1}^I$ in \mathcal{Q} ,

$$\max_{1 \leq i \leq Q} (\mathbf{x}_i - \mathbf{y}_i)^T (\mathbf{F}_i(\mathbf{x}) - \mathbf{F}_i(\mathbf{y})) \geq c_{\text{uP}} \|\mathbf{x} - \mathbf{y}\|^2. \quad (19)$$

The constant c_{uP} is called uniformly P constant.

If a function \mathbf{F} enjoys one of the properties above, we will also say that the corresponding VI $(\mathcal{Q}, \mathbf{F})$ enjoys the property (i.e., if \mathbf{F} is monotone, we say that the VI is monotone, etc...).

Note that in the case of affine functions, $\mathbf{F}(\mathbf{x}) = \mathbf{M}\mathbf{x} + \mathbf{b}$, there is no difference between strict monotonicity and strongly monotonicity, and the uniform P property coincides with the P property. Monotonicity properties play in the VI realm the same role that convex functions play in optimization. In fact, we recall that a differentiable function f is convex, strictly, strongly convex on a convex set \mathcal{Q} if and only if its gradient is monotone, strictly, strongly monotone on \mathcal{Q} . The P properties can be viewed as an extension of the monotonicity properties tailored to the possible partitioned structure of the VI; when the partitioned VI has only one block, i.e., $I = 1$, the P properties collapse to the corresponding monotonicity properties. In Figure 1 we summarize in a pictorial way some well established relations between these various classes along with some of their consequences. Theorem 3 provides instead a formal statement of some existence and uniqueness results that will be used throughout the paper.

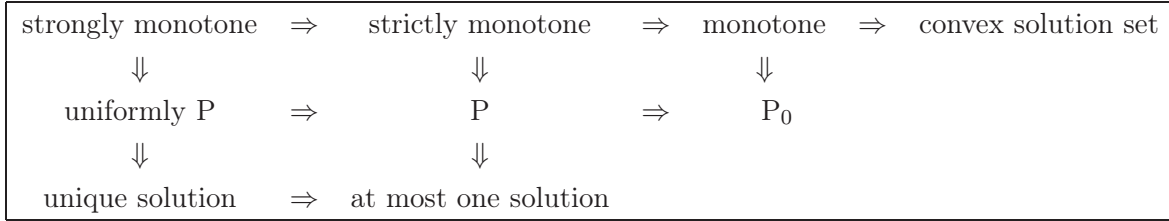


Figure 1: Monotonicity and their consequences on VIs.

Theorem 3 *Given the VI $(\mathcal{Q}, \mathbf{F})$, suppose that \mathcal{Q} is closed and convex and \mathbf{F} is continuous on \mathcal{Q} . The following statements hold:*

- (a) *The VI $(\mathcal{Q}, \mathbf{F})$ has a (possibly empty) closed solution set. If \mathcal{Q} is bounded, the solution set is nonempty and thus compact [30, Cor. 2.2.5];*
- (b) *If \mathbf{F} is monotone on \mathcal{Q} , then the VI $(\mathcal{Q}, \mathbf{F})$ has a (possibly empty) convex solution set [30, Th. 2.3.5];*
- (c) *If \mathbf{F} is strictly monotone on \mathcal{Q} , then the VI $(\mathcal{Q}, \mathbf{F})$ has at most one solution [30, Th. 2.3.3(a)]; the same conclusion holds if the VI $(\mathcal{Q}, \mathbf{F})$ is partitioned and \mathbf{F} is a P function on \mathcal{Q} [30, Prop. 3.5.10(a)];*
- (d) *If \mathbf{F} is strongly monotone on \mathcal{Q} , then the VI $(\mathcal{Q}, \mathbf{F})$ has a unique solution [30, Th. 2.3.3(b)]; the same conclusion holds if the VI $(\mathcal{Q}, \mathbf{F})$ is partitioned and \mathbf{F} is a uniformly P function on \mathcal{Q} [30, Prop. 3.5.10(b)].*

Note that the uniqueness results stated in part (c)-(d) do not require that the set \mathcal{Q} be bounded. For P_0 partitioned VIs more refined existence/uniqueness results can be obtained; we report them in Appendix A.

3.2 Conditions for the monotonicity and P properties

The results in the previous subsection and many of the algorithmic developments to follow hinge critically on the function \mathbf{F} enjoying some monotonicity or P property. However, using directly the definition to establish whether such properties hold is in general not possible. It is then useful to be able to establish more practical conditions to check the aforementioned properties. It is well known that when \mathcal{Q} is an *open* set and \mathbf{F} is continuously differentiable on \mathcal{Q} , with Jacobian matrix denoted by \mathbf{JF} , it holds that [30, Prop. 2.3.2]:²

$$\begin{aligned} \mathbf{F}(\mathbf{x}) \text{ is monotone on } \mathcal{Q} & \Leftrightarrow \mathbf{JF}(\mathbf{x}) \succeq \mathbf{0}, \forall \mathbf{x} \in \mathcal{Q}; \\ \mathbf{F}(\mathbf{x}) \text{ is strictly monotone on } \mathcal{Q} & \Leftarrow \mathbf{JF}(\mathbf{x}) \succ \mathbf{0}, \forall \mathbf{x} \in \mathcal{Q}; \\ \mathbf{F}(\mathbf{x}) \text{ is strongly monotone on } \mathcal{Q} & \Leftrightarrow \mathbf{JF} - c_{\text{sm}} \mathbf{I} \succeq \mathbf{0}, \forall \mathbf{x} \in \mathcal{Q}; \end{aligned} \quad (20)$$

where $\mathbf{A} \succeq \mathbf{0}$ ($\mathbf{A} \succ \mathbf{0}$) means that \mathbf{A} is a positive semidefinite (definite) matrix. The verification of these kind of conditions is often difficult and, furthermore, in many practical instances their verification cannot easily be linked to physical characteristics of the systems being studied. Therefore, our aim in this subsection is that of developing some (conceptually) simpler conditions that permit to deduce the desired \mathbf{F} properties and that, at least in some instances, can give some further insight into the problem at hand. It is interesting to note that, as we will see later on, the conditions we introduce here on pure theoretical grounds also have some important algorithmic consequences.

Let us define the matrix \mathbf{JF}_{low} having the same dimension as $\mathbf{JF}(\mathbf{x})$:

$$[\mathbf{JF}_{\text{low}}]_{rs} \triangleq \begin{cases} \inf_{\mathbf{x} \in \mathcal{Q}} [\mathbf{B}^T \mathbf{JF}(\mathbf{x}) \mathbf{B}]_{rr}, & \text{if } r = s, \\ -\sup_{\mathbf{x} \in \mathcal{Q}} |[\mathbf{B}^T \mathbf{JF}(\mathbf{x}) \mathbf{B}]_{rs}|, & \text{otherwise,} \end{cases} \quad (21)$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is an arbitrary nonsingular matrix; and, under the assumption that \mathbf{F} and \mathcal{Q} have a partitioned structure, let us also introduce the “condensed” $I \times I$ real matrices $\Upsilon_{\mathbf{F}}$ and $\Gamma_{\mathbf{F}}$:

$$[\Upsilon_{\mathbf{F}}]_{ij} \triangleq \begin{cases} \alpha_i^{\min}, & \text{if } i = j, \\ -\beta_{ij}^{\max}, & \text{otherwise,} \end{cases} \quad \text{and} \quad [\Gamma_{\mathbf{F}}]_{ij} \triangleq \begin{cases} 0, & \text{if } i = j, \\ \frac{\beta_{ij}^{\max}}{\alpha_i^{\min}}, & \text{otherwise,} \end{cases} \quad (22)$$

with

$$\alpha_i^{\min} \triangleq \inf_{\mathbf{x} \in \mathcal{Q}} \lambda_{\text{least}}(\mathbf{C}_i^T \mathbf{J}_i \mathbf{F}_i(\mathbf{x}) \mathbf{C}_i) \quad \text{and} \quad \beta_{ij}^{\max} \triangleq \sup_{\mathbf{x} \in \mathcal{Q}} \|\mathbf{C}_i^T \mathbf{J}_j \mathbf{F}_i(\mathbf{x}) \mathbf{C}_j\|, \quad (23)$$

where $\lambda_{\text{least}}(\mathbf{A})$ denotes the smallest eigenvalue of the symmetric part of matrix \mathbf{A} , $\mathbf{J}_j \mathbf{F}_i(\mathbf{x})$ is the Jacobian of $\mathbf{F}_i(\mathbf{x})$ with respect to \mathbf{x}_j , and $\mathbf{C}_i \in \mathbb{R}^{n_i \times n_i}$ with $i = 1, \dots, I$, is a set of arbitrary nonsingular matrices. Note that in the definition of $\Gamma_{\mathbf{F}}$ we tacitly assumed all $\alpha_i^{\min} \neq 0$ and β_{ij}^{\max} are finite; the latter condition is equivalent to the boundedness of $\mathbf{J}_j \mathbf{F}_i(\mathbf{x})$ on \mathcal{Q} . Matrices \mathbf{B} and \mathbf{C}_i ’s provide an additional degree of freedom in obtaining conditions for monotonicity and P properties of \mathbf{F} that can be linked to physical characteristics of the systems being studied (see Sec. 7 for some examples). In order to explore the relationship between the two matrices $\Upsilon_{\mathbf{F}}$ and $\Gamma_{\mathbf{F}}$, we need the following definition.

²Conditions in (20) can be generalized also to the case in which \mathcal{Q} is closed; this will be done in Sec. 6, where we introduce the VI problem in the complex domain; see Proposition 29.

Definition 4 A matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ is called P matrix if every principal minor of \mathbf{M} is positive.

Many equivalent characterizations for a P matrix can be given. The interested reader is referred to [51, 52] for details; here we recall only the following properties, instrumental for our further derivations. Any positive definite matrix is a P-matrix, but the reverse does not hold (unless the matrix is symmetric). Furthermore, building on the properties of the P-matrices [51, Lemma 13.14], one can show that $\Upsilon_{\mathbf{F}}$ is a P-matrix if and only if $\rho(\Gamma_{\mathbf{F}}) < 1$, where $\rho(\mathbf{A})$ denotes the spectral radius of \mathbf{A} .

Matrices \mathbf{JF}_{low} and $\Upsilon_{\mathbf{F}}$ are useful to obtain sufficient conditions for the monotonicity and P property of the mapping \mathbf{F} , as given next.

Proposition 5 Let $\mathbf{F} : \mathcal{Q} \rightarrow \mathbb{R}^n$ be continuously differentiable with bounded derivatives on the closed and convex set \mathcal{Q} . The following statements hold:

- (a) If \mathbf{JF}_{low} is copositive,³ then \mathbf{F} is monotone on \mathcal{Q} ;
- (b) If \mathbf{JF}_{low} is strictly copositive,² then \mathbf{F} is strictly monotone on \mathcal{Q} ;
- (c) If \mathbf{JF}_{low} is positive definite, then \mathbf{F} is strongly monotone on \mathcal{Q} with strong monotonicity constant given by $c_{\text{sm}} = \lambda_{\text{least}}(\mathbf{JF}_{\text{low}})$ [or $c_{\text{sm}} = \lambda_{\text{least}}(\Upsilon_{\mathbf{F}})$].⁴

If we assume a Cartesian product structure, i.e. $\mathbf{F} = (\mathbf{F}_i(\mathbf{x}))_{i=1}^I$ and $\mathcal{Q} = \prod_i \mathcal{Q}_i$, then:

- (d) If $\Upsilon_{\mathbf{F}}$ is positive semidefinite/ P_0 -matrix, then \mathbf{F} is a monotone/ P_0 function on \mathcal{Q} ;
- (e) If $\Upsilon_{\mathbf{F}}$ is a P-matrix [which is equivalent to $\rho(\Gamma_{\mathbf{F}}) < 1$], then \mathbf{F} is a uniformly P-function on \mathcal{Q} with uniform P constant given by

$$\hat{c}_{\text{uP}}(\mathbf{F}) = \frac{\delta(\Upsilon_{\mathbf{F}})/Q}{(1 + \zeta(\Upsilon_{\mathbf{F}})/\delta(\Upsilon_{\mathbf{F}}))^{2(Q-1)}}, \quad (24)$$

where $\zeta(\Upsilon_{\mathbf{F}}) \triangleq \max_{r \neq q} |[\Upsilon_{\mathbf{F}}]_{rq}|$, and $\delta(\Upsilon_{\mathbf{F}}) \triangleq \min\{\sigma([\Upsilon_{\mathbf{F}}]_{\alpha\alpha}) : \alpha \subseteq \{1, \dots, I\}\}$, with $\sigma([\mathbf{M}]_{\alpha\alpha})$ denoting the smallest of the real eigenvalues (if any exists) of the principal submatrix of \mathbf{M} of order α .

Remark 6 (On the P-properties) The P property as stated in the proposition above is of a block kind (\mathbf{F} is partitioned in component blocks of dimension $n_i \geq 1$), as opposed to the point kind (\mathbf{F} is partitioned in component blocks of dimension $n_i = 1$), as analyzed in [30, Ch. 3]. Furthermore, because of this block P property, Proposition 5 does not follow from [30, Proposition 3.5.9] that pertains to the point P property. A special case of Proposition 5(e) can be found in [31], where the block P property is stated for a mapping \mathbf{F} associated to some NEPs in the context of CR systems. \square

³A matrix \mathbf{A} is copositive if $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$; it is strictly copositive if $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{0} \neq \mathbf{x} \geq \mathbf{0}$. A positive (semi)definite matrix is (strictly) copositive.

⁴The least eigenvalue of a real (nonnecessarily symmetric) matrix \mathbf{A} , denoted by $\lambda_{\text{least}}(\mathbf{A})$, is the smallest eigenvalue of the symmetric part of \mathbf{A} .

4 Nash Equilibrium Problems: Fundamentals

In a standard real NEP there are I players each controlling a variable $\mathbf{x}_i \in \mathbb{R}^{n_i}$ that must belong to the player's feasible set \mathcal{Q}_i , which is assumed to be closed and convex: $\mathbf{x}_i \in \mathcal{Q}_i$. In what follows we denote by $\mathbf{x} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_I)$, the vector of all players' variables, while $\mathbf{x}_{-i} \triangleq (\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_I)$ denote the vector of all players' strategies variables except that of player i . The aim of player i , given the other players' strategies \mathbf{x}_{-i} , is to choose an $\mathbf{x}_i \in \mathcal{Q}_i$ that minimizes his cost function $f_i(\mathbf{x}_i, \mathbf{x}_{-i})$, i.e.,

$$\begin{aligned} & \underset{\mathbf{x}_i}{\text{minimize}} && f_i(\mathbf{x}_i, \mathbf{x}_{-i}) \\ & \text{subject to} && \mathbf{x}_i \in \mathcal{Q}_i. \end{aligned} \tag{25}$$

Note that the players' optimization problem are *coupled* since the players' objective function (may) depend on the other players' choices. Define the joint strategy set of the NEP by $\mathcal{Q} = \prod_{i=1}^I \mathcal{Q}_i$, whereas $\mathcal{Q}_{-i} \triangleq \prod_{j \neq i} \mathcal{Q}_j$, and set $\mathbf{f} \triangleq (f_i)_{i=1}^I$. The NEP is formally defined by the tuple $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$. A solution of the NEP is the well-known Nash Equilibrium (NE), which is formally defined in (2).

A useful way to see a NE is as a fixed-point of the best-response mapping for each player; this suggests the use of (iterative) best-response-based algorithms to solve the game. Given the limitations of classical fixed-point results in the study of convergence of best-response based algorithms (cf. Sec. 1), we address this issue by reducing a NEP to a VI problem, and then using the well developed theory of VIs [30]. This approach has the advantage of leading quite naturally to the derivation of implementable distributed algorithms along with their convergence properties.

4.1 Connection to variational inequalities

At the basis of the VI approach to NEPs there is an easy equivalence between a real NEP and a suitably defined partitioned VI. This equivalence follows readily from the minimum principle for convex problems and the Cartesian structure of the joint strategy set \mathcal{Q} [30, Prop. 1.4.2].

Proposition 7 *Given the real NEP $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$, suppose that for each player i the following hold:*

- i) *the (nonempty) strategy set \mathcal{Q}_i is closed and convex;*
- ii) *the payoff function $f_i(\mathbf{x}_i, \mathbf{x}_{-i})$ is convex and continuously differentiable in \mathbf{x}_i for every fixed \mathbf{x}_{-i} .*

Then, the game \mathcal{G} is equivalent to the VI(\mathcal{Q}, \mathbf{F}), where $\mathbf{F}(\mathbf{x}) \triangleq (\nabla_{\mathbf{x}_i} f_i(\mathbf{x}))_{i=1}^I$.

In the sequel we refer to the VI(\mathcal{Q}, \mathbf{F}) defined in previous proposition as the VI associated to the NEP \mathcal{G} . It is possible to relax the assumptions in Proposition 7 and still get useful connections between games and VIs [20]; but since our aims are mainly computational, we don't pursue this topic further. Indeed, throughout the paper, we will make the following blanket convexity/smoothness assumptions, unless stated otherwise.

Assumption 1. For each $i = 1, \dots, I$, the set \mathcal{Q}_i is a nonempty, closed, and convex subset of \mathbb{R}^{n_i} and the function $f_i(\mathbf{x}_i, \mathbf{x}_{-i})$ is continuously differentiable on $\mathcal{Q} = \prod_i \mathcal{Q}_i$ and convex in \mathbf{x}_i for every fixed $\mathbf{x}_{-i} \in \mathcal{Q}_{-i}$.

Assumption 2. For each $i = 1, \dots, I$, each function $f_i(\mathbf{x})$ is twice continuously differentiable with bounded derivatives on $\mathcal{Q} = \prod_i \mathcal{Q}_i$.

4.2 Existence and uniqueness of a NE

Building on the VI reformulation in the previous section and the existence/uniqueness results for partitioned VIs given in Theorem 3, we can easily state the following theorem that needs no further proof.

Theorem 8 *Given the real NEP $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$, suppose that \mathcal{G} satisfies Assumption 1 and let $\mathbf{F}(\mathbf{x}) \triangleq (\nabla_{\mathbf{x}_i} f_i(\mathbf{x}))_{i=1}^I$. Then, the following statements hold:*

- (a) *Suppose that for every i the strategy set \mathcal{Q}_i is bounded. Then the NEP has a nonempty and compact solution set;*
- (b) *Suppose that $\mathbf{F}(\mathbf{x})$ is a monotone function on \mathcal{Q} . Then the NEP has a convex (possibly empty) solution set;*
- (c) *Suppose that $\mathbf{F}(\mathbf{x})$ is a P (or strictly monotone) function on \mathcal{Q} . Then the NEP has at most one solution;*
- (d) *Suppose that $\mathbf{F}(\mathbf{x})$ is a uniformly-P (or strongly monotone) function on \mathcal{Q} . Then the NEP has a unique solution.*

Sufficient conditions for $\mathbf{F}(\mathbf{x})$ being a strictly, strongly monotone or a (uniformly-)P function on \mathcal{Q} are given in Proposition 5. The condition given in Theorem 8(a) guaranteeing the existence of a solution for a NEP requires that the strategy set of each player be bounded. More general results on the existence of a solution for a NEP with unbounded strategy sets can be found in Appendix A. Note that the (existence and) uniqueness results stated in parts (b)-(d) do not require that the set \mathcal{Q} be bounded.

Remark 9 (On the uniqueness conditions) According to Proposition 5(b), and supposing that $\mathbf{F}(\mathbf{x}) = (\nabla_{\mathbf{x}_i} f_i(\mathbf{x}))_{i=1}^I$ is continuously differentiable with bounded derivatives on \mathcal{Q} (Assumption 2), a sufficient condition for the uniqueness of the NE is that the matrix $\Upsilon_{\mathbf{F}}$ defined in (22) is a P matrix. It turns out that this condition is sufficient also for global convergence of best-response asynchronous distributed algorithms described in Sec. 5. Note that if $\Upsilon_{\mathbf{F}}$ is a P matrix, it must be $\alpha_i^{\min} = \inf_{\mathbf{z} \in \mathcal{Q}} [\lambda_{\min}(\nabla_{\mathbf{x}_i}^2 f_i(\mathbf{z}))] > 0$ for all i , where $\lambda_{\min}(\nabla_{\mathbf{x}_i}^2 f_i(\mathbf{z}))$ denotes the minimum eigenvalue of $\nabla_{\mathbf{x}_i}^2 f_i(\mathbf{z})$. Thus an implicit consequence of the P assumption on the matrix $\Upsilon_{\mathbf{F}}$ is the uniform positive definiteness of the matrices $\nabla_{\mathbf{x}_i}^2 f_i$ on \mathcal{Q} , which implies the uniformly strong convexity of $f_i(\cdot, \mathbf{x}_{-i})$ for any given $\mathbf{x}_{-i} \in \mathcal{Q}_{-i}$ and thus the uniqueness of the solution of the i -th player's optimization problem, for any given $\mathbf{x}_{-i} \in \mathcal{Q}_{-i}$. This latter implication is a quite natural consequence of the uniqueness of the solution of the NE: one cannot expect the game to have only one NE if the players' problems have multiple solutions.

The observations above also pave the way for a nice interpretation of the P condition. Note first that the β 's in the definition of the matrix $\Upsilon_{\mathbf{F}}$ measure the coupling of the players' optimization problems: the larger the β 's, the more coupled the players' subproblems are. Indeed, if all the β 's were 0, the game $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ would decompose into I uncoupled optimization problems; in such a case, requiring the matrix $\Upsilon_{\mathbf{F}}$ to be P simply amounts to requiring all α 's to be positive, which obviously implies uniqueness of the solution. It is reasonable that if the β 's increase from zero but remain small enough with respect to the α 's, the game will still have a unique solution. The P property quantifies how large the β 's can grow while still preserving the uniqueness of the solution. \square

We conclude this subsection providing a sufficient condition for the matrix $\Upsilon_{\mathbf{F}}$ in (22) to be a P (positive definite) matrix, which can be derived by elementary diagonal dominance arguments.

Proposition 10 *The matrix $\Upsilon_{\mathbf{F}}$ in (22) is a P-matrix if one of the following two sets of conditions are satisfied: for some $\mathbf{w} = (w_i)_{i=1}^I > \mathbf{0}$,*

$$\frac{1}{w_i} \sum_{j \neq i} w_j \frac{\beta_{ij}^{\max}}{\alpha_i^{\min}} < 1, \quad \forall i = 1, \dots, I, \quad \frac{1}{w_j} \sum_{i \neq j} w_i \frac{\beta_{ij}^{\max}}{\alpha_j^{\min}} < 1, \quad \forall j = 1, \dots, I. \quad (26)$$

If actually both conditions in (26) are satisfied, then $\Upsilon_{\mathbf{F}}$ is positive definite.

The sufficient conditions in Proposition 10 will be shown in Sec. 7 to have an interesting physical interpretation in the context of power control problems in CR systems.

4.3 Problem classes

Based on the previous results, it is natural to introduce the following classes of real NEPs, which will play a fundamental role in the rest of the paper.

Definition 11 *A real NEP $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ is:*

- i) *a monotone Nash equilibrium problem if Assumption 1 holds and the associated $VI(\mathcal{Q}, \mathbf{F})$ is monotone;*
- ii) *a uniformly P Nash equilibrium problem if Assumption 1 holds and the associated $VI(\mathcal{Q}, \mathbf{F})$ is uniformly P;*
- iii) *a P_{Υ} Nash equilibrium problem if Assumptions 1 and 2 hold and the matrix $\Upsilon_{\mathbf{F}}$ of the associated $VI(\mathcal{Q}, \mathbf{F})$ is P.*

From the development so far, Figure 2 summarizing the relations between these classes of problems needs no further explanation, except for the fact that P_{Υ} (and thus uniformly P) NEPs are not a subclass of monotone NEPs. This is shown by the following example.

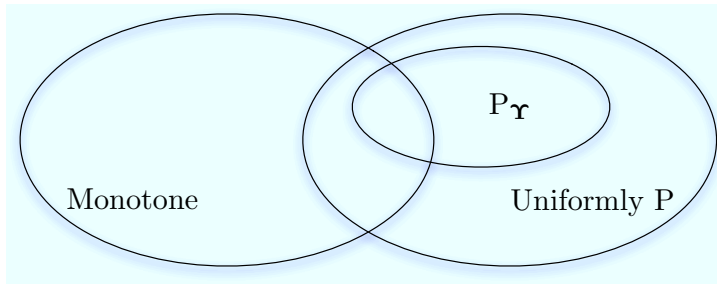


Figure 2: Relation among NEP classes.

Example 12 (A P_{Υ} NEP which is not monotone) Consider a real NEP with two players, each controlling one scalar variable: x_1 and x_2 . The players' problems are

$$\begin{array}{ll} \text{minimize}_{x_1} & \frac{1}{2}x_1^2 + 4x_1x_2 \\ \text{subject to} & x_1 \in [0, 10] \end{array} \quad \begin{array}{ll} \text{minimize}_{x_2} & \frac{1}{2}x_2^2 - \frac{1}{8}x_1x_2 \\ \text{subject to} & x_2 \in [-2, 2] \end{array}$$

The VI associated to this is $\text{VI}([0, 10] \times [-2, 2], \mathbf{F})$, with $\mathbf{F} = [x_1 + 4x_2, x_2 - (1/8)x_1]^T$. The symmetric part of \mathbf{JF} , \mathbf{JF}_s , and the matrix $\Upsilon_{\mathbf{F}}$ are given by:

$$\mathbf{JF}_s = \begin{bmatrix} 1 & 31/16 \\ 31/16 & 1 \end{bmatrix}, \quad \Upsilon_{\mathbf{F}} = \begin{bmatrix} 1 & -4 \\ -1/8 & 1 \end{bmatrix}.$$

Since \mathbf{JF}_s has a negative determinant, \mathbf{F} cannot be monotone; on the other hand it is easy to check that the two principal minors of $\Upsilon_{\mathbf{F}}$ are positive, implying that \mathcal{G} is a P NEP.

By Theorem 8, a uniformly P (and thus a P_{Υ}) NEP has a unique solution, whereas monotone NEPs have a convex solution set that can be empty. While obviously a NEP with a unique solution is not necessarily a P NEP, the latter class of problems covers an interesting array of problems, as we will be shown later on.

Centralized algorithms for monotone and uniformly P_{Υ} NEPs, based on VI theory, are well-known [30, vol II]; in this paper, we focus on the more challenging issue of devising *distributed* (and possibly asynchronous) solution schemes for NEPs, which is the topic of the next section.

5 Distributed Algorithms for NEPs

This section along with the next one constitute the core theoretical part of the paper. We develop here a novel theory that allows devising *distributed* algorithms for computing Nash equilibria in several significant settings. More specifically, we will provide novel distributed (asynchronous) algorithms for the solution of

- (a) monotone NEPs;
- (b) P_{Υ} NEPs.

Since monotone NEPs may have multiple solutions, in case (a), we will further consider both the situations in which one is interested in computing *any one* solution, and the situations in which one wants to select the *best* solution, according to a given criterion. Settings (a) and (b) cover a wide range of situations and they will be shown in Sec. 7 to encompass several important equilibrium models in signal processing and communications; in particular the models proposed in Sec. 2. In each of the settings above we will provide best-response-based distributed algorithms along with their convergence properties; the proposed algorithms differ in: i) the computational effort; ii) the players' synchronization/signaling requirements; and iii) the convergence speed. Note that while centralized solution methods are known for uniformly P NEPs, the development of distributed algorithms for this class of games is at the time of this writing an open problem.

This section is organized in three parts. Sections 5.1 and 5.2 focus on algorithms for P_{Υ} and monotone NEPs, respectively; results in this sections will be the building blocks for the more difficult issue of equilibrium selection problem addressed in Sec. 5.3.

5.1 Best-response distributed algorithms for P_{Υ} NEPs

Since in a NEP every player is trying to minimize his own objective function, a natural approach to compute a solution of a NEP is to consider an iterative algorithm wherein all the players, given the strategies of the others and according to a given scheduling (e.g., sequentially or simultaneously), update their own strategy

by solving their optimization problem (25). Here, we focus on a very general class of best-response-based algorithms, namely the *totally asynchronous* best-response algorithms (in the sense specified in [53]). In these schemes, some players may update their strategies more frequently than others and they may even use an outdated information about the strategy profile used by the others; which is very appealing in many practical multiuser communication systems, such as wireless ad-hoc networks or CR systems wherein synchronization requirements are hard to enforce.

To provide a formal description of the algorithm, we need to introduce some preliminary definitions. In an asynchronous scheme, the users may not update their own strategies at each iteration; let denote then by $\mathcal{T}_i \subseteq \mathcal{T} \subseteq \{0, 1, 2, \dots\}$ the set of times at which player i updates his own strategy \mathbf{x}_i , denoted by $\mathbf{x}_i^{(n)}$ (thus, implying that, at time $n \notin \mathcal{T}_i$, $\mathbf{x}_i^{(n)}$ is left unchanged). Moreover, in computing their optimal strategy, the users can use an outdated version of the others' strategies; let then $\tau_j^i(n)$ be the most recent time at which the strategy profile of player j is perceived by player i at the n -th iteration (observe that $\tau_j^i(n)$ satisfies $0 \leq \tau_j^i(n) \leq n$). Hence, if player i updates its strategy at the n -th iteration, then he minimizes his cost function using the following (possibly) outdated strategy profile of the other players:

$$\mathbf{x}_{-i}^{(\tau^i(n))} \triangleq \left(\mathbf{x}_1^{(\tau_1^i(n))}, \dots, \mathbf{x}_{i-1}^{(\tau_{i-1}^i(n))}, \mathbf{x}_{i+1}^{(\tau_{i+1}^i(n))}, \dots, \mathbf{x}_I^{(\tau_I^i(n))} \right). \quad (27)$$

Some standard conditions in asynchronous convergence theory, which are fulfilled in any practical implementation, need to be satisfied by the schedule \mathcal{T}_i 's and $\tau_j^i(n)$'s, namely for each i :

- A1)** $0 \leq \tau_j^i(n) \leq n$ (at any given iteration n , each player i can use only the strategy profile $\mathbf{x}_{-i}^{(\tau^i(n))}$ adopted by the other players in the previous iterations);
- A2)** $\lim_{k \rightarrow \infty} \tau_j^i(n_k) = +\infty$, where $\{n_k\}$ is a sequence of elements in \mathcal{T}_i that tends to infinity (for any given iteration index n_k , the values of the components of $\mathbf{x}_{-i}^{(\tau^i(n))}$ in (27) generated prior to n_k are not used in the updates of $\mathbf{x}_i^{(n)}$, when n becomes sufficiently larger than n_k);
- A3)** $|\mathcal{T}_i| = \infty$ (no player fails to update his own strategy as time n goes on).

Using the above definitions, the totally asynchronous algorithm based on the best-responses of the players is described in Algorithm 1. The convergence properties of the algorithm are given in Theorem 13.

Algorithm 1: Asynchronous Best-Response Algorithm

(S.0) : Choose any feasible $\mathbf{x}^{(0)} \in \mathcal{Q}$ and set $n = 0$.

(S.1) : If $\mathbf{x}^{(n)}$ satisfies a suitable termination criterion: **STOP**

(S.2) : **for** $i = 1, \dots, I$, compute

$$\mathbf{x}_i^{(n+1)} = \begin{cases} \mathbf{x}_i^* \in \underset{\mathbf{x}_i \in \mathcal{Q}_i}{\operatorname{argmin}} f_i \left(\mathbf{x}_i, \mathbf{x}_{-i}^{(\tau^i(n))} \right), & \text{if } n \in \mathcal{T}_i \\ \mathbf{x}_i^{(n)}, & \text{otherwise} \end{cases} \quad (28)$$

(S.3) : $n \leftarrow n + 1$; go to (S.1).

Theorem 13 Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a P_{Υ} NEP. Any sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ generated by Algorithm 1 converges to the unique NE of \mathcal{G} , for any given updating schedule of the players satisfying assumptions A1-A3.

Proof. See Appendix B. ■

Remark 14 (Flexibility of the algorithm) Algorithm 1 contains as special cases a large number of algorithms, each one obtained by a possible choice of the schedule of the users in the updating procedure (i.e., the parameters $\{\tau_j^i(n)\}$ and $\{\mathcal{T}_i\}$). Examples are the *simultaneous* (Jacobi scheme) and *sequential* (Gauss-Seidel scheme) updates, where the players update their own strategies *simultaneously* and *sequentially*, respectively. Indeed, the Jacobi update corresponds to the schedule $\tau_j^i(n) = n$ and $\mathcal{T}_i = \{1, 2, \dots\}$ for all i and j , whereas the Gauss-Seidel scheme is obtained by taking $\tau_j^i(n) = n$ and $\mathcal{T}_i = \{i, i + I, i + 2I, \dots\}$ for all i and j . Moreover, variations of such a totally asynchronous scheme, e.g., including constraints on the maximum tolerable delay in the updating and on the use of the outdated information (which leads to the so-called *partially asynchronous algorithms*), can also be considered [53]. An important result stated in Theorem 13 is that all the algorithms resulting as special cases of Algorithm 1 are guaranteed to reach the unique NE of the NEP, under the same set of convergence conditions, since the matrix $\Upsilon_{\mathbf{F}}$ does not depend on the particular choice of $\{\tau_j^i(n)\}$ and $\{\mathcal{T}_i\}$. Note that all the algorithms coming from Algorithm 1 are robust against missing or outdated updates of the players. This feature strongly relaxes the constraints on the synchronization of the players' updates; which makes this class of algorithms appealing in many practical distributed systems.

Note that the (synchronous) projection-response algorithms for monotone VIs (and thus NEPs) proposed in [34] and [30, 33] are not guaranteed to converge if applied to a P_{Υ} NEP that is not monotone. □

Remark 15 (On the convergence conditions of best-response algorithms) Global convergence of Algorithm 1 is guaranteed under the P property of $\Upsilon_{\mathbf{F}}$ (or equivalently $\rho(\Gamma_{\mathbf{F}}) < 1$). We have already pointed out however that such a condition cannot be satisfied if there is a player whose cost function has a singular Hessian, even in just one point. In fact, if this is the case, we have, $\alpha_i^{\min} = 0$ for some i , let us say $i = 1$ without loss of generality, which implies that the matrix $\Gamma_{\mathbf{F}}$ has a 1 in the left-upper corner. Since the matrix $\Gamma_{\mathbf{F}}$ is nonnegative, we have that this implies $\rho(\Gamma_{\mathbf{F}}) \geq 1$ [54, Th. 1.7..4]. Assuming that the element 1 is contained in an irreducible principal matrix, we will actually have $\rho(\Gamma_{\mathbf{F}}) > 1$. Note that the irreducibility assumption is extremely weak and trivially satisfied if the matrix $\Gamma_{\mathbf{F}}$ is positive, which is true in many applications. In the next section we discuss a remedy for this issue. □

5.2 Proximal distributed algorithms for monotone NEPs

In this section we deal with monotone NEPs (see Definition 11). The main difference between P_{Υ} (or uniformly P) NEPs and monotone NEPs is that the former always have a unique solution, while monotone NEPs may have multiple solutions, and the players' objective functions need not be uniformly strongly convex; see Remark 15. It turns out that, for such games, Algorithm 1 may fail to converge. There is a host of solution methods available in the literature to solve monotone real VIs and thus monotone NEPs (see, e.g., [30, Vol. II]), but these algorithms are centralized. Recently, in [34], the authors proposed some distributed synchronous schemes for solving monotone VIs, based on the gradient-response mapping; we have already discussed the main drawbacks of these algorithms, see Sec. 1 (see also Sec. 7 for some numerical results).

The development of distributed *best-response* algorithms for solving monotone NEPs with (possibly) multiple solutions is a challenging task; in this subsection, we provide a first answer to this issue building on a regularization technique known as proximal algorithms, see [30, Ch 12] for an introduction to proximal point methods for VIs. The proposed approach is to reduce the solution of a *single* monotone NEP to the solution of a *sequence* of P_{Υ} NEPs with a particular structure. The advantage of this method is that we can efficiently solve each of the P_{Υ} NEPs with convergence guarantee using Algorithm 1 introduced in the previous section; the disadvantage is that, to recover the solution of the original single monotone NEP, we have to solve a (possibly infinite) number of P_{Υ} NEPs. However, it is important to remark from the outset that this potential drawback is greatly mitigated by the fact that, as we discuss shortly, (i) we only need to solve these P_{Υ} NEPs inaccurately; (ii) the (inaccurate) solution of the P_{Υ} NEPs usually requires little computational effort; and (iii) in practice, a fairly accurate solution of the original NEP is obtained after the solution of a limited number of P_{Υ} NEPs.

Before introducing the formal description of the algorithm, let us begin with some simple observations motivating how the sequence of P_{Υ} NEPs is built. Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP; consider a perturbation of this game defined as $\mathcal{G}_{\tau, \mathbf{y}} = \langle \mathcal{Q}, (f_i + (\tau/2) \|\bullet - \mathbf{y}_i\|^2)_{i=1}^I \rangle$, where τ is a positive parameter and $\mathbf{y} = (\mathbf{y}_i)_{i=1}^I$ is a given vector in \mathbb{R}^n , with each $\mathbf{y}_i \in \mathbb{R}^{n_i}$; we term \mathbf{y} center of the regularization. Note that $\mathcal{G}_{\tau, \mathbf{y}}$ is the game wherein each player i , anticipating $\mathbf{x}_{-i} \in \mathcal{Q}_{-i}$, solves the following convex optimization problem:

$$\begin{aligned} & \underset{\mathbf{x}_i}{\text{minimize}} && f_i(\mathbf{x}_i, \mathbf{x}_{-i}) + \frac{\tau}{2} \|\mathbf{x}_i - \mathbf{y}_i\|^2 \\ & \text{subject to} && \mathbf{x}_i \in \mathcal{Q}_i. \end{aligned} \tag{29}$$

Let us consider now the VI reformulations of \mathcal{G} and $\mathcal{G}_{\tau, \mathbf{y}}$, given by $\text{VI}(\mathcal{Q}, \mathbf{F})$ and $\text{VI}(\mathcal{Q}, \mathbf{F}_{\tau, \mathbf{y}})$ respectively, where $\mathbf{F}(\mathbf{x}) \triangleq (\nabla_{\mathbf{x}_i} f_i(\mathbf{x}))_{i=1}^I$ and $\mathbf{F}_{\tau, \mathbf{y}}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \tau(\mathbf{x} - \mathbf{y})$, and let us introduce the matrices $\Upsilon_{\mathbf{F}}$ and $\Upsilon_{\mathbf{F}_{\tau, \mathbf{y}}}$ associated to \mathcal{G} and $\mathcal{G}_{\tau, \mathbf{y}}$, respectively. It is not difficult to check that

$$\Upsilon_{\mathbf{F}_{\tau, \mathbf{y}}} = \Upsilon_{\mathbf{F}} + \tau \mathbf{I}. \tag{30}$$

Note that $\Upsilon_{\mathbf{F}_{\tau, \mathbf{y}}}$ does not depend on \mathbf{y} . It follows readily from (30) that if τ is large enough, $\Upsilon_{\mathbf{F}_{\tau, \mathbf{y}}}$ is a P matrix, meaning that $\mathcal{G}_{\tau, \mathbf{y}}$ is a P_{Υ} NEP, for any given $\mathbf{y} \in \mathbb{R}^n$. More specifically, using the definitions of β_{ij}^{\max} 's and α_i^{\min} 's as given in (23), we have the following.

Lemma 16 *Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP. For any given $\mathbf{y} \in \mathbb{R}^n$, the game $\mathcal{G}_{\tau, \mathbf{y}} = \langle \mathcal{Q}, (f_i + (\tau/2) \|\bullet - \mathbf{y}_i\|^2)_{i=1}^I \rangle$ is a P_{Υ} NEP for every τ larger than $\bar{\tau}$ (independent of \mathbf{y}), with*

$$\bar{\tau} \triangleq \max_{1 \leq i \leq I} \left\{ \sum_{j \neq i} \beta_{ij}^{\max} - \alpha_i^{\min} \right\}. \tag{31}$$

Although the assumption that \mathcal{G} be monotone will play a key role in the forthcoming derivations, it is important to remark that what we need for the above lemma to hold is only that the cost functions f_i 's in the game have bounded second order derivatives on \mathcal{Q} .

Nice as it is, the result above would be of no practical interest if we were not able to connect the solutions of $\mathcal{G}_{\tau, \mathbf{y}}$ to those of \mathcal{G} . Indeed, the solutions of \mathcal{G} and $\mathcal{G}_{\tau, \mathbf{y}}$ are in general different but, nevertheless, there exists a connection between them: a point \mathbf{x}^* is a solution of \mathcal{G} if and only if \mathbf{x}^* is a solution of $\mathcal{G}_{\tau, \mathbf{x}^*}$.

Proposition 17 Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP. For any given $\tau > 0$, $\mathbf{x}^* \in \mathcal{Q}$ is a solution of \mathcal{G} if and only if \mathbf{x}^* is a solution of $\mathcal{G}_{\tau, \mathbf{x}^*}$.

Proof. See Appendix C. ■

Lemma 16 and Proposition 17 open the way to the design of convergent distributed algorithms for monotone NEPs, as shown next. Let us choose τ being large enough so that $\mathcal{G}_{\tau, \mathbf{y}}$ is a P_{Υ} NEP (cf. Lemma 16). It follows from Theorem 8 that $\mathcal{G}_{\tau, \mathbf{y}}$ has a unique solution, denoted by $\mathbf{S}_{\tau}(\mathbf{y})$. Using $\mathbf{S}_{\tau}(\mathbf{y})$, Proposition 17 can be restated as follows: \mathbf{x}^* is a solution of \mathcal{G} if and only if it is a fixed point of $\mathbf{S}_{\tau}(\bullet)$, i.e., $\mathbf{x}^* = \mathbf{S}_{\tau}(\mathbf{x}^*)$. It seems then natural to compute the solutions of \mathcal{G} using the fixed-point-type iteration $\mathbf{x}^{(n+1)} = \mathbf{S}_{\tau}(\mathbf{x}^{(n)})$, starting from a feasible point $\mathbf{x}^{(0)}$; which corresponds to solving the sequence of NEPs $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$ for $n = 0, 1, \dots$. If τ is sufficiently large [e.g., as in (31)], each $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$ is a P_{Υ} NEP (cf. Lemma 16), and thus its unique solution can be computed in a distributed way with convergence guarantee by the asynchronous best-response algorithm described in Algorithm 1 (cf. Theorem 13). The above discussion motivates the following algorithm for computing the solutions of a monotone NEP, whose convergence properties are given in Theorem 18 below.

Algorithm 2: Proximal Decomposition Algorithm (PDA)

Data : Let $\tau > 0$ be given.

(S.0) : Choose any feasible $\mathbf{x}^{(0)} \in \mathcal{Q}$ and set $n = 0$.

(S.1) : If $\mathbf{x}^{(n)}$ satisfies a suitable termination criterion: STOP.

(S.2) : Solve the game $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$ and set $\mathbf{x}^{(n+1)} \triangleq \mathbf{S}_{\tau}(\mathbf{x}^{(n)})$

(S.3) : $n \leftarrow n + 1$; go to (S.1).

Theorem 18 Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP with a nonempty solution set. Suppose that τ is large enough so that $\Upsilon_{\mathbf{F}_{\tau, \mathbf{y}}}$ is a P matrix. Then, Algorithm 2 is well defined, and the sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ generated by the algorithm converges to a solution of the game \mathcal{G} .

Algorithm 2 is of great conceptual interest, but its applicability is limited, unless one is able to easily compute $\mathbf{S}_{\tau}(\mathbf{x}^{(n)})$. Although there are interesting problems in which this can be done efficiently (see Sec. 7), in general one is expected to solve a number of P_{Υ} NEPs, each of them requiring an infinite iterative method to compute each $\mathbf{S}_{\tau}(\mathbf{x}^{(n)})$. To overcome this issue, we propose next a variant of Algorithm 2, in which suitable approximations of $\mathbf{S}_{\tau}(\mathbf{x}^{(n)})$ can be used. Algorithm 3 below describes such a variant, where we have added a further degree of freedom in the updating rule: the new iteration $\mathbf{x}^{(n+1)}$ is not necessarily given by (an approximation of) $\mathbf{S}_{\tau}(\mathbf{x}^{(n)})$, but lies instead on the line connecting the old iteration $\mathbf{x}^{(n)}$ to (the approximation of) $\mathbf{S}_{\tau}(\mathbf{x}^{(n)})$.

Algorithm 3: Approximate Proximal Decomposition Algorithm (APDA)

Data : Let $\{\varepsilon^{(n)}\}_{n=0}^{\infty}$, $\{\eta^{(n)}\}_{n=0}^{\infty}$ and $\tau > 0$ be given.

(S.0) : Choose any feasible $\mathbf{x}^{(0)} \in \mathcal{Q}$ and set $n = 0$.

(S.1) : If $\mathbf{x}^{(n)}$ satisfies a suitable termination criterion: STOP.

(S.2) : Solve the game $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$ within the accuracy $\varepsilon^{(n)}$: Find a $\mathbf{z}^{(n)}$ s.t. $\|\mathbf{z}^{(n)} - \mathbf{S}_{\tau}(\mathbf{x}^{(n)})\| \leq \varepsilon^{(n)}$.

(S.3) : Set $\mathbf{x}^{(n+1)} \triangleq (1 - \eta^{(n)})\mathbf{x}^{(n)} + \eta^{(n)}\mathbf{z}^{(n)}$.

(S.4) : $n \leftarrow n + 1$; go to (S.1).

The error term $\varepsilon^{(n)}$ measures the accuracy used at iteration n in computing the solution $\mathbf{S}_\tau(\mathbf{x}^{(n)})$ of $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$. The parameter $\eta^{(n)}$ instead, introduces a memory in the updating rule, it establishes where exactly we move along the line passing through the old iterations $\mathbf{x}^{(n)}$ and $\mathbf{z}^{(n)}$. Note that if we take $\varepsilon^{(n)} = 0$ and $\eta^{(n)} = 1$ for all n , Algorithm 3 reduces to Algorithm 2. The advantage of Algorithm 3 with respect to Algorithm 2 is that $\mathbf{z}^{(n)}$ can be computed in a finite number of steps, so that Algorithm 3 becomes implementable in practice. Obviously, the errors $\varepsilon^{(n)}$'s and the parameters $\eta^{(n)}$'s must be chosen according to some suitable conditions, if one wants to guarantee convergence. These conditions are established in the following theorem.

Theorem 19 *Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP with a nonempty solution set. Suppose that τ is large enough so that $\mathbf{\Upsilon}_{\mathbf{F}_{\tau, \mathbf{y}}}$ is a P-matrix. Choose $\{\varepsilon^{(n)}\} \subset [0, \infty)$ such that $\sum_{n=1}^{\infty} \varepsilon^{(n)} < \infty$ and $\{\eta^{(n)}\} \subset [R_m, R_M]$, with $0 < R_m \leq R_M < 2$. Then, Algorithm 3 is well defined, and the sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ generated by the algorithm converges to a solution of \mathcal{G} .*

The proof of Theorem 19 (and thus also Theorem 18) is a consequence of the following facts and thus is omitted: i) [30, Th. 12..3.9]; ii) The observation that \mathcal{G} is equivalent to the VI(\mathcal{Q}, \mathbf{F}) (Proposition 7), with $\mathbf{F} = (\nabla_{\mathbf{x}_i} f_i)_{i=1}^I$, and the VI(\mathcal{Q}, \mathbf{F}) has a solution; and iii) Under the P property of $\mathbf{\Upsilon}_{\mathbf{F}_{\tau, \mathbf{y}}}$, Step S.2 of Algorithm 3 (and Algorithm 2) is well defined (Theorem 13).

It is interesting to remark that for sake of simplicity we assumed τ to be a fixed number. However, τ can be varied from iteration to iteration provided that $\tau \in (\bar{\tau}, \tau^{\max}]$, where τ^{\max} is any finite number.

While the utility of having the possibility to use inexact solutions in Step S.2 of Algorithm 3 is apparent, the usefulness of Step S.3 is less evident. This kind of “averaging” is known as over-relaxation and has its roots in classical successive over-relaxation methods for solving systems of linear equations [55, Sec. 7.4]. In our context, the extra degree of freedom offered by Step S.3 can bring numerical improvements; see, e.g., [30].

Algorithm 3 is conceptually a double loop scheme wherein at each (outer) iteration n , given $\mathbf{x}^{(n)}$, one needs to compute the approximation $\mathbf{z}^{(n)}$, which requires an inner iterative process. Since the condition $\sum_{n=1}^{\infty} \varepsilon^{(n)} < \infty$ implies $\varepsilon^{(n)} \downarrow 0$, when the iterations progress, $\mathbf{S}_\tau(\mathbf{x}^{(n)})$ has to be estimated with an increasing accuracy. However, in practice, this is not a problem since when iterations progress, $\{\mathbf{x}^{(n)}\}$ usually converges, implying $\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\| \rightarrow 0$. One can then use $\mathbf{x}^{(n)}$ as a good approximation to initialize any (inner) procedure in Step 2 to compute $\mathbf{z}^{(n)}$. It turns out that, in spite of the increasing precision requirements, in practice a suitable $\mathbf{z}^{(n)}$ in Step 2 can be computed very easily.

Finally, observe that a natural choice for computing $\mathbf{z}^{(n)}$ in Step 2 of Algorithm 3 is Algorithm 1. When this choice is made, Algorithm 3 also becomes an asynchronous method, having all the desired features described in the previous section. The only difference with Algorithm 1 is that, in Algorithm 3, “from time to time” (more precisely when the inner termination test $\|\mathbf{z}^{(n)} - \mathbf{S}_\tau(\mathbf{x}^{(n)})\| \leq \varepsilon^{(n)}$ in Step 2 is satisfied) the objective function of the players are changed by updating the regularizing term from $\frac{\tau}{2} \|\mathbf{x}_i - \mathbf{x}_i^{(n)}\|^2$ to $\frac{\tau}{2} \|\mathbf{x}_i - \mathbf{x}_i^{(n+1)}\|^2$, which generally requires some coordination among the players to establish when a satisfactory approximation $\mathbf{z}^{(n)}$ has been reached. The remark below deals exactly with issues related to this aspect.

Remark 20 (On the Inner Termination Criterion) We have seen that, in Step 2 of Algorithm 3, the players must be able to decide whether $\|\mathbf{z}^{(n)} - \mathbf{S}_\tau(\mathbf{x}^{(n)})\| \leq \varepsilon^{(n)}$ holds. In the following we suggest a simple distributed protocol to do that. Observe preliminarily that an error bound on the distance of the current

iteration $\mathbf{z}^{(n)}$ from the solution $\mathbf{S}_\tau(\mathbf{x}^{(n)})$ of $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$ can be obtained by solving a convex (quadratic) problem (see, e.g., [30, Prop. 6.3.1], [30, Prop. 6.3.7]). For example, under the P properties of $\Upsilon_{\mathbf{F}_{\tau, \mathbf{x}^{(n)}}}$ defined in (30), the following error bound holds for the game $\mathcal{G}_{\tau, \mathbf{x}^{(n)}}$ [30, Prop. 6.3.1]: a (finite) constant $c > 0$ exists such that

$$\|\mathbf{z} - \mathbf{S}_\tau(\mathbf{x}^{(n)})\| \leq c \|\mathbf{F}_\tau^{\text{nat}}(\mathbf{z})\| \quad \forall \mathbf{z} \in \mathcal{Q}, \quad (32)$$

where $\mathbf{F}_\tau^{\text{nat}}(\mathbf{z}) \triangleq \mathbf{z} - \Pi_{\mathcal{Q}}(\mathbf{z} - \mathbf{F}(\mathbf{z}) - \tau \mathbf{z})$, with $\Pi_{\mathcal{Q}}(\mathbf{x})$ denoting the Euclidean projection of \mathbf{x} onto the closed and convex set \mathcal{Q} . Note that, since \mathcal{Q} has a Cartesian structure, $\mathbf{F}_\tau^{\text{nat}}(\mathbf{z})$ can be partitioned as $\mathbf{F}_\tau^{\text{nat}}(\mathbf{z}) = ([\mathbf{F}_\tau^{\text{nat}}(\mathbf{z})]_i)_{i=1}^I$, where each $[\mathbf{F}_\tau^{\text{nat}}(\mathbf{z})]_i = \mathbf{z}_i - \Pi_{\mathcal{Q}_i}(\mathbf{z}_i - \mathbf{F}_i(\mathbf{z}) - \tau \mathbf{z}_i)$ can be locally computed by the associated player i by solving a quadratic programming, as long as $\mathbf{F}_i(\mathbf{z})$ is available at the player side.⁵

Using (32), the implementation of the Step 2 of Algorithm 3 can be obtained as follows. Each player i chooses preliminarily a suitable local termination sequence $\{\varepsilon_i^{(n)}\}_n \subset [0, \infty)$ such that $\sum_{n=1}^{\infty} \varepsilon_i^{(n)} < \infty$; the termination criterion of each player i becomes then $\|[\mathbf{F}_\tau^{\text{nat}}(\mathbf{x}^{(n)})]_i\| \leq \varepsilon_i^{(n)}$, which can be locally implemented. Once the desired local accuracy is reached by all the players, they can all update the center of their regularization. Note that this protocol guarantees that the requirement on the sequence $\varepsilon^{(n)}$ in Step 2 as stated in Theorem 19 is met, since $\varepsilon^{(n)} \triangleq \sum_{i=1}^I \varepsilon_i^{(n)}$ satisfies $\sum_{n=1}^{\infty} \varepsilon^{(n)} < \infty$. The last issue to address for a practical implementation of this protocol is to understand how the players can detect the others having reached the desired termination criterion. This can be done by exchanging one bit of information, if some signaling is allowed; otherwise each user can just update its regularization after experiencing no changes in $\|[\mathbf{F}_\tau^{\text{nat}}(\mathbf{x}^{(n)})]_i\|$ (or his cost function) for a prescribed number of iterations. \square

5.3 Equilibrium selection for monotone NEPs

In the previous section we discussed distributed algorithms for the computation of a solution of monotone NEPs. A feature of these algorithms is that they converge under mild conditions that do not imply the uniqueness of the NE of the NEP. In the presence of multiple equilibria, however, the proposed algorithms do not allow to perform any selection of the solution they reach, but they may converge in principle to *any* NE of the game; which makes the achievable system performance unpredictable. It would be interesting instead to be able to select, among all the solutions of the game, the one(s) that satisfies some additional criterion. We refer to this problem as *equilibrium selection problem*. In this section, we address this issue; the outcome will be a novel set of distributed algorithms along with their convergence properties that solve the equilibrium selection problem; this additional feature comes at the price of a (moderate) increase of the complexity in computing the players' best-response solution and signaling among the players.

Let us introduce first an informal description of the algorithm. Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP and let $\text{SOL}(\mathcal{Q}, \mathbf{f})$ denote its solution set, assumed to be nonempty without loss of generality. Recall that since \mathcal{G} is monotone, $\text{SOL}(\mathcal{Q}, \mathbf{f})$ is always convex (cf. Theorem 8). Stated in mathematical terms, the equilibrium selection problem consists in solving the following bi-level optimization problem:

⁵In many practical applications, as those considered in the second part of the paper, each $[\mathbf{F}_\tau^{\text{nat}}(\mathbf{z})]_i$ can be computed by local measurements from the players. For instance, this happens when the strategies of the other players affect the cost function of each player by a term that is locally measurable as additive interference noise.

$$\begin{aligned}
& \underset{\mathbf{x}}{\text{minimize}} && \phi(\mathbf{x}) \\
& \text{subject to} && \mathbf{x} \in \text{SOL}(\mathcal{Q}, \mathbf{f}),
\end{aligned} \tag{33}$$

where the function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is assumed to be continuously differentiable and convex. The function ϕ thus defines the additional criterion according to which one wants to select a solution in the set of the Nash equilibria of the game \mathcal{G} : solving (33) indeed corresponds to choosing the NE of \mathcal{G} that minimizes ϕ . Note that, under the monotonicity of \mathcal{G} , (33) is a convex optimization problem [$\text{SOL}(\mathcal{Q}, \mathbf{f})$ is convex]. However, standard solution techniques cannot be applied because the feasible set $\text{SOL}(\mathcal{K}, \mathbf{f})$ is only implicitly defined and, in general, it is not expressed as a standard system of inequalities. To overcome this difficulty, and in the same spirit of the previous section, instead of attacking problem (33) directly, we propose to solve a sequence of standard regularized NEPs (“standard” means a game whose players’ feasible sets are not of an implicit type, and therefore can be solved by classic methods, like Algorithm 1). Each standard regularized game has the following structure $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}} = \langle \mathcal{Q}, (f_i + \varepsilon \phi + (\tau/2) \|\bullet - \mathbf{y}_i\|^2)_{i=1}^I \rangle$, where ε and τ are fixed positive constants and $\mathbf{y} \triangleq (\mathbf{y}_i)_{i=1}^I$ is a given point in \mathbb{R}^n with each $\mathbf{y}_i \in \mathbb{R}^{n_i}$; $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}}$ is a NEP wherein each player i , anticipating $\mathbf{x}_{-i} \in \mathcal{Q}_{-i}$, solves the following convex optimization problem:

$$\begin{aligned}
& \underset{\mathbf{x}_i}{\text{minimize}} && f_i(\mathbf{x}_i, \mathbf{x}_{-i}) + \varepsilon \phi(\mathbf{x}_i, \mathbf{x}_{-i}) + \frac{\tau}{2} \|\mathbf{x}_i - \mathbf{y}_i\|^2 \\
& \text{subject to} && \mathbf{x}_i \in \mathcal{Q}_i.
\end{aligned} \tag{34}$$

Note that the players’ problems in this game differ from (29) in the presence of the additional term $\varepsilon \phi(\mathbf{x}_i, \mathbf{x}_{-i})$ in the objective function. It is not surprising that the function ϕ appears in the objective function of the players; roughly speaking, it represents the additional amount of information to be included in the game to “drive” the system toward the desired solution.

Proceeding as in the previous section, we can now establish the connection between the regularized NEPs $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}}$ and the equilibrium selection problem (33). First of all note that, in the setting of problem (33), the NEPs $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}}$ is equivalent to the VI($\mathcal{Q}, \mathbf{F}_{\tau, \varepsilon, \mathbf{y}}$) where $\mathbf{F}_{\tau, \varepsilon, \mathbf{y}} \triangleq \mathbf{F} + \varepsilon \nabla \phi + \tau (\mathbf{I} - \mathbf{y})$ and $\mathbf{I} : \mathbf{x} \mapsto \mathbf{x}$ is the identity map. Denoting by $\mathbf{\Upsilon}_{\mathbf{F}_{\tau, \varepsilon, \mathbf{y}}}$ the matrix defined in (22) and associated to $\mathbf{F}_{\tau, \varepsilon, \mathbf{y}}$, Lemma 21 below shows that there exists a sufficiently large τ such that $\mathbf{\Upsilon}_{\mathbf{F}_{\tau, \varepsilon, \mathbf{y}}}$ is a P matrix, implying that $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}}$ is a $P_{\mathbf{\Upsilon}}$ NEP.

Lemma 21 *Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP; let $\phi : \mathbb{R}^n \mapsto \mathbb{R}$ be a continuously differentiable function on \mathcal{Q} whose gradient $\nabla \phi$ is Lipschitz continuous on \mathcal{Q} , with constant L_ϕ ; and let $\bar{\varepsilon} > 0$ be given. For any fixed $\mathbf{y} \in \mathbb{R}^n$, the game $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}} = \langle \mathcal{Q}, (f_i + \varepsilon \phi + (\tau/2) \|\bullet - \mathbf{y}_i\|^2)_{i=1}^I \rangle$ with $\varepsilon \in [0, \bar{\varepsilon}]$ is a $P_{\mathbf{\Upsilon}}$ NEP for every τ larger than $\bar{\tau}_{\bar{\varepsilon}}$ (independent on \mathbf{y} and ε)*

$$\bar{\tau}_{\bar{\varepsilon}} \triangleq \max_{1 \leq i \leq I} \left\{ \sum_{j \neq i} \beta_{ij}^{\max} - \alpha_i^{\min} \right\} + (I - 1) \bar{\varepsilon} L_\phi, \tag{35}$$

where β_{ij}^{\max} ’s and α_i^{\min} ’s as defined in (23).

Under the setting of Lemma 21, the game $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}}$ is a $P_{\mathbf{\Upsilon}}$ NEP and thus has a unique solution, denoted by $\mathbf{S}_{\tau, \varepsilon}(\mathbf{y})$ [cf. Theorem 8]; such a unique solution $\mathbf{S}_{\tau, \varepsilon}(\mathbf{y})$ can be computed with convergence guaranteed

using Algorithm 1 on $\mathcal{G}_{\tau, \varepsilon, \mathbf{y}}$. The solution of the original equilibrium selection problem (33) can be recovered using the fixed-point-type iteration $\mathbf{x}^{(n+1)} = \mathbf{S}_{\tau, \varepsilon(n)}(\mathbf{x}^{(n)})$, starting from a feasible point $\mathbf{x}^{(0)}$ and by suitably varying $\varepsilon(n)$; which corresponds to solve the sequence of NEPs $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$ for $n = 0, 1, \dots$. This procedure is made formal in Algorithm 4 below.

Algorithm 4: Proximal-Tikhonov Regularization Algorithm (PTRA)

Data : Let $\{\varepsilon^{(n)}\} \downarrow 0$ and $\tau > 0$ be given.

(S.0) : Choose any feasible $\mathbf{x}^{(0)} \in \mathcal{Q}$ and set $n = 0$.

(S.1) : If $\mathbf{x}^{(n)}$ satisfies a suitable termination criterion, STOP.

(S.2) : Set $\mathbf{x}^{(n+1)}$ to be the solution of the game $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$.

(S.3) : Set $n \leftarrow n + 1$ and return to (S.1).

Note that, since in Algorithm 4 the sequence $\{\varepsilon^{(n)}\}$ converges to zero, there always exists an $\bar{\varepsilon} > 0$ such that $\varepsilon^{(n)} \in [0, \bar{\varepsilon}]$, implying by Lemma 21 that for a sufficiently large τ all the games $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$ in the Step 2 of the algorithm are P_{Υ} NEPs and thus have a unique solution; this makes the sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ generated by Algorithm 4 well defined. The convergence properties of the algorithm are given in the following theorem.

Theorem 22 *Let $\mathcal{G} = \langle \mathcal{Q}, \mathbf{f} \rangle$ be a monotone NEP with a nonempty solution set $\text{SOL}(\mathcal{Q}, \mathbf{f})$. Consider the equilibrium selection problem (33) and suppose that i) ϕ is continuously differentiable and convex on \mathcal{Q} ; ii) the level sets of ϕ on $\text{SOL}(\mathcal{Q}, \mathbf{f})$ are bounded; and iii) $\nabla \phi$ is Lipschitz continuous on \mathcal{Q} , with constant L_{ϕ} . Moreover, suppose that τ is large enough so that $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$ is a P_{Υ} NEP for any n , and choose the sequence $\{\varepsilon^{(n)}\}$ such that $\varepsilon^{(n)} > 0$ for all n , $\{\varepsilon^{(n)}\} \downarrow 0$, and $\sum_{n=0}^{\infty} \varepsilon^{(n)} = \infty$. Then Algorithm 4 is well defined; the sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ is bounded; and every of its limit points is a solution of (33).*

Proof. See Appendix D. ■

Theorem 22 guarantees convergence of the algorithm under mild assumptions. Conditions on ϕ are pretty standard; in particular, assumptions i) and ii) together with the monotonicity of \mathcal{G} state that the optimization problem (33) is convex and admits a solution (for the sake of simplicity we also assumed differentiability of ϕ); whereas iii) guarantees that there exists a finite (large enough) τ such that $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$ is a P_{Υ} NEP (cf. Lemma 21). The assumptions on the sequence $\{\varepsilon^{(n)}\}_{n=1}^{\infty}$ are also rather weak and require $\varepsilon^{(n)}$ to go to zero, but not too fast; which can be satisfied, e.g., by taking $\varepsilon^{(n)} = 1/(1 + na)$, with $n = 0, 1, 2, \dots$ and a being any positive constant. The reason for this requirement is rather intuitive: if $\varepsilon^{(n)}$ became “too small, too soon”, then the term $\frac{\tau}{2} \|\mathbf{x}_i - \mathbf{y}_i\|^2$ in the players’ objective functions would dominate the term $\varepsilon^{(n)} \phi$, making the role of $\varepsilon^{(n)} \phi$ negligible; Algorithm 4 would become “almost” Algorithm 3, so that one could not guarantee to find nothing more than a point in $\text{SOL}(\mathcal{Q}, \mathbf{f})$.

The implementation of Algorithm 4 requires the ability of solving at each round n the P_{Υ} NEP $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$. To this end, as already observed in the previous section, we can use several centralized methods. However, if we are interested in distributed solution schemes, Algorithm 1 applied to $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$ is the natural choice. Note that the convergence conditions of the algorithm applied to $\mathcal{G}_{\tau, \varepsilon(n), \mathbf{x}^{(n)}}$ as stated in Theorem 13 are always met, provided τ is large enough; see, e.g., (35) in Lemma 21.

As in the previous section, note that, unless one has simple ways to compute the solutions of the games $\mathcal{G}_{\tau, \varepsilon^{(n)}, \mathbf{x}^{(n)}}$, Algorithm 4 requires at each step the exact computation of the solution of the regularized games $\mathcal{G}_{\tau, \varepsilon^{(n)}, \mathbf{x}^{(n)}}$ (inner loop), which in principle requires a conceptually infinite procedure. While in practice this might not be a problem, it still leaves open the question of whether, paralleling the results of the previous section, one can develop versions of Algorithm 4 where inexact solutions are used in the Step 2. The answer to this question is positive, but the corresponding theory is rather complex and, for sake of simplicity, we prefer to omit it here; the interest reader can work it out using results in [21].

5.4 A bird's-eye view

In the previous three sections we proposed several distributed algorithms, which are applicable to different scenarios. It is useful to summarize the results obtained so far, showing that, in spite of apparent diversities, all the algorithms belong to a same family; Figure 3 provides the roadmap of the proposed distributed solution methods for real NE

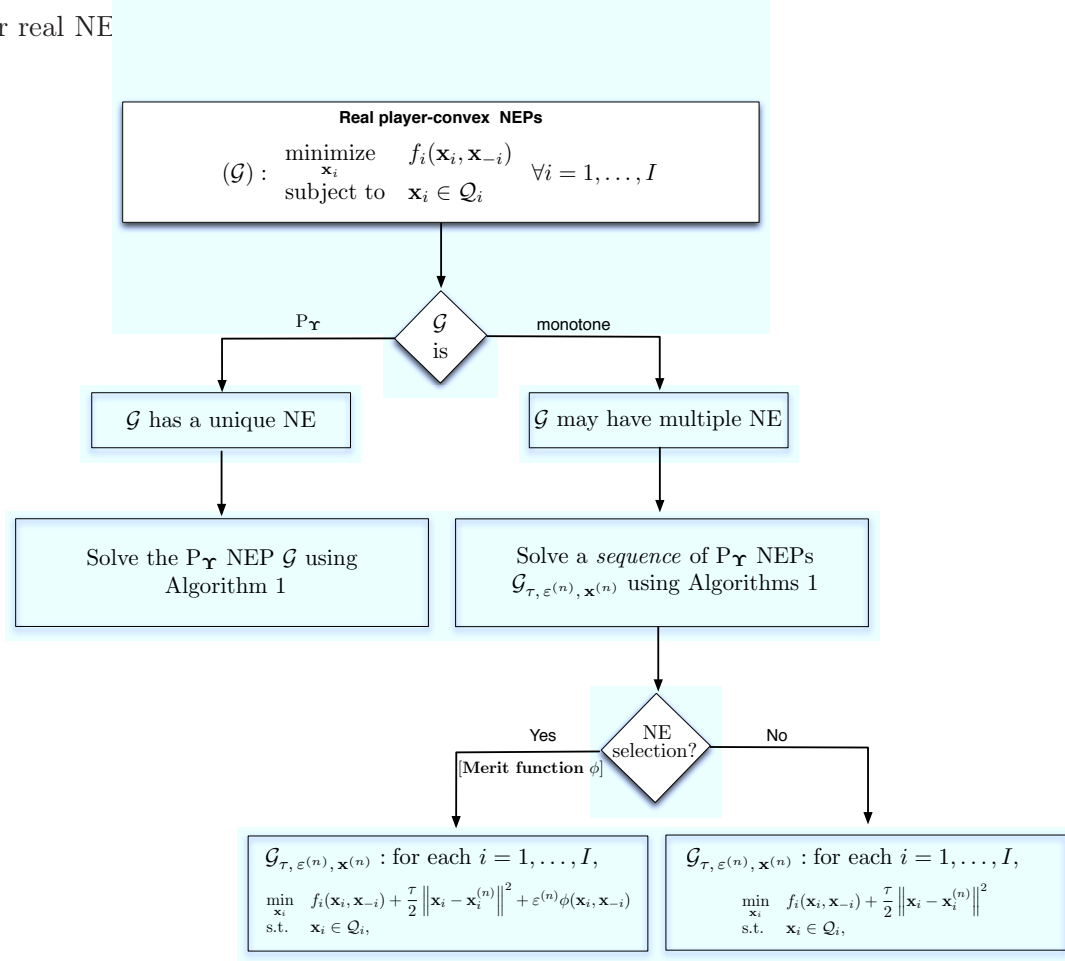


Figure 3: The roadmap of the proposed distributed solution methods for real player-convex NEPs

Conceptually, what we have proposed is indeed a unified algorithm, where the users can explicitly choose the degree of desired cooperation and signaling, converging to solutions having different performance, namely: i) *any one* NE, when there is no cooperation among users, and ii) the “best” NE (according to an outer merit

function ϕ), when there is some cooperation. The choice of one scheme in favor to the other as well as the merit function ϕ will depend then on the trade-off between signaling and performance that the users are willing to exchange/achieve. The core of the proposed solution methods can be summarized in the following unified updating rule: at iteration n , the optimal strategy of each user i is

$$\mathbf{x}_i^{(n+1)} = \underset{\mathbf{x}_i \in \mathcal{Q}_i}{\operatorname{argmin}} \left\{ f_i(\mathbf{x}_i, \mathbf{x}_{-i}^{(n)}) + \pi_i^{(\nu_n)}(\mathbf{x}_i, \mathbf{x}_{-i}^{(n)}) + \frac{\tau}{2} \left\| \mathbf{x}_i - \mathbf{x}_i^{(\nu_n)} \right\|^2 \right\} \quad (36)$$

where the first term $f_i(\mathbf{x}_i, \mathbf{x}_{-i}^{(n)})$ is the usual term in an iterative best-response algorithm, the second term $\pi_i^{(\nu_n)}(\mathbf{x}_i, \mathbf{x}_{-i}^{(n)})$ (whose update is performed at iteration ν_n) can be interpreted as a nonlinear pricing in the objective function of the users, and the third term $\frac{\tau}{2} \left\| \mathbf{x}_i - \mathbf{x}_i^{(\nu_n)} \right\|^2$ is a (proximal) regularization. Observe that there are two iteration indexes: n is the main discrete-time unit, whereas ν_n is increased every few discrete-time units (e.g., if $\nu_n = \lfloor n/10 \rfloor$, then ν_n is updated every 10 discrete-time units).

The price function $\pi_i^{(\nu_n)}$ can be interpreted as a measure of the “altruism/selfishness” of the users and represents the trade-off factor between signaling and performance. Indeed, we may have the following:

- $\pi_i^{(\nu_n)} = 0$ (no cooperation): The users are not willing to cooperate; the best one can get is converge to *any one* solution of the game (i.e., with no control on the quality of the solution); this is guaranteed *even in the presence of multiple equilibria* if the NEP is monotone (P_{Υ});
- $\pi_i^{(\nu_n)} \neq 0$ (some cooperation): The users may exchange some signaling in the form of pricing through the function $\pi_i^{(\nu_n)}(\mathbf{x}_i, \mathbf{x}_{-i}^{(n)}) = \varepsilon^{(\nu_n)} \phi(\mathbf{x}_i, \mathbf{x}_{-i}^{(n)})$ and converge to the NE that minimizes the merit function $\phi(\mathbf{x})$; convergence is guaranteed if the NEP is monotone.

Remark 23 (Role of pricing) It is important to remark that the pricing term $\pi_i^{(\nu_n)}$ does not need to be linear; moreover it has a specific and well understood role in the optimization of the system performance. This is a fundamental departure from current literature that uses *linear* pricing as a heuristic to improve the performance of a NE in power control games (see, e.g., [56] for scalar power control problems); in these works there is neither a proof of convergence of the modified game nor a theoretical explanation of the performance improvement due to pricing. As a direct product of our framework, we obtain instead a clear understanding of the meaning of the pricing; for example, a linear price in the form $\varepsilon^{(\nu_n)} \boldsymbol{\pi}_i^T \mathbf{x}_i$, with $\varepsilon^{(\nu_n)} \rightarrow 0$, corresponds to the selection of a NE that minimizes the linear function $\sum_i \boldsymbol{\pi}_i^T \mathbf{x}_i$, resulting likely in better system performance.

Finally, the proximal regularization in (36) has the role to numerically “stabilize” the algorithm; the proximal gain τ is the trade-off factor between the convergence stability and the convergence speed. The beneficial effects of a proximal regularization are well understood in the optimization literature; see, e.g., [53].

6 Variational Inequalities and Games in the Complex Domain

All the results presented so far apply to real NEPs. However, in many applications, e.g., in digital communications, array processing, and signal processing, the variables involved in the optimization are complex numbers. For instance, in the MIMO problems introduced in Sec. 2, the optimization variables of the players

are complex matrices. For these applications, the reformulation of the problem into the real domain is awkward, and generally leads to final conditions that cannot be easily rewritten in terms of the original complex setup. It seems instead more convenient to work directly in the complex domain. This naturally calls for the definition of the VI problem in the domain of complex matrices, which is the aim of this section. The novel contribution in this direction is threefold, as listed next.

Following the same approach as in the real case (cf. Sec. 3), we introduce in Sec. 6.2 the minimum principle for constrained convex optimization problems in the domain of complex matrices, generalizing the already known complex gradient-vanishing conditions obtained in [40] for the *unconstrained* minimization of real-valued functions of complex variables. As an intermediate result, we also introduce a Taylor expansion of real-valued functions of complex matrices that is amenable to our MIMO applications; details can be found in Appendix E. The second important contribution is given in Sec. 6.3, where we provide the formulation of a VI problem in the domain of complex variables that can be seen as the natural generalization of the minimum principle introduced in Sec. 6.2, where the (complex matrix) gradient is replaced by a (complex-valued matrix) function of matrix variables. We then introduce the concepts of monotonicity and P properties for the complex-valued matrix function of the VI and provide matrix conditions for this properties to hold. Interestingly, these conditions are the natural generalization of those obtained in Sec. 3.1, provided that a new definition of Jacobian matrix tailored to complex-valued matrix functions is introduced. Finally, the third contribution is the connection between VIs and NEPs in the complex domain, as given in Sec. 6.5. Our derivations are based on the use of \mathbb{R} -matrix derivatives (also termed Wirtinger derivatives) that are preliminary introduced in Sec. 6.1; this section is motivated by the lack of a well-established notation and definitions for \mathbb{R} -matrix derivatives of functions of complex (matrix) variables; two good tutorials on the subject (termed $\mathbb{C}\mathbb{R}$ -calculus) are [40, 57] (note that [40] and [57] use a different formalism, we follow [40]).

6.1 \mathbb{R} -matrix derivatives

In practical applications, we often deal with optimization of real-valued functions $f : \mathbb{C} \ni z \mapsto f(z) \in \mathbb{R}$ of a complex variable z that are not differentiable in \mathbb{C} (termed also \mathbb{C} -differentiable or holomorphic).⁶ However, the same univariate function $f : \mathbb{C} \rightarrow \mathbb{R}$ can also be viewed as a bivariate function of its real and imaginary components, i.e., $f(z) = g(z_R, z_I)$, where $g : \mathbb{R}^2 \mapsto \mathbb{R}$ is a real-valued function of the real variables $z_R \triangleq \text{Re}(z)$ and $z_I \triangleq \text{Im}(z)$. This way, one may be able to replace the nonexistence of the \mathbb{C} -derivative of f with the existence of the real partial derivatives of $g(z_R, z_I)$, which is actually what one needs to compute a stationary point of the function. This motivates the introduction of the so-called \mathbb{R} -derivative and conjugate \mathbb{R} -derivative of $f : \mathbb{C} \rightarrow \mathbb{R}$ at $z_0 \in \mathbb{C}$, formally defined as

$$\frac{\partial f}{\partial z}(z_0) \triangleq \frac{1}{2} \left(\frac{\partial f(z)}{\partial z_R} - j \frac{\partial f(z)}{\partial z_I} \right) \Big|_{z=z_0} \quad \text{and} \quad \frac{\partial f}{\partial z^*}(z_0) \triangleq \frac{1}{2} \left(\frac{\partial f(z)}{\partial z_R} + j \frac{\partial f(z)}{\partial z_I} \right) \Big|_{z=z_0}, \quad (37)$$

respectively, where $j = \sqrt{-1}$. Note that the derivatives above must be interpreted formally, because z and its conjugate z^* in (37) are treated as they were mutually independent; the derivatives $\frac{\partial f}{\partial z_R}$ and $\frac{\partial f}{\partial z_I}$ represent instead the true (non-formal) partial derivatives of f viewed as a bivariate function of z_R and z_I , i.e., $f = \tilde{f}(z_R, z_I)$. When $\frac{\partial f}{\partial z_R}$ and $\frac{\partial f}{\partial z_I}$ exist (and are continuous), implying that (37) is well-defined, we say

⁶It is a known fact that nonconstant *real-valued* functions (of complex variables) are not \mathbb{C} -differentiable.

that f is \mathbb{R} -differentiable (or continuously \mathbb{R} -differentiable); similarly to the real case, when we say that a function $f : \mathcal{K} \rightarrow \mathbb{R}$ is \mathbb{R} -differentiable (or continuously \mathbb{R} -differentiable) on the closed set \mathcal{K} , we mean that the function is so on an open set containing \mathcal{K} .

The \mathbb{R} -derivatives defined in (37) for a real-valued function can be naturally extended to *complex-valued* functions of a complex argument, that is, $f : \mathbb{C} \rightarrow \mathbb{C}$; formally we still have (37), but now $f(z) = \check{f}(z_R, z_I) \triangleq \check{f}_R(z_R, z_I) + j \cdot \check{f}_I(z_R, z_I)$, with $\check{f} : \mathbb{R}^2 \mapsto \mathbb{C}$ and $\check{f}_R, \check{f}_I : \mathbb{R}^2 \mapsto \mathbb{R}$, and by $\partial f / \partial z_R$ we mean $\partial \check{f}_R / \partial z_R + j \cdot \partial \check{f}_I / \partial z_R$ (similarly for $\partial f / \partial z_I$).

When f is a (complex-valued) scalar function of complex *matrices*, that is $f : \mathbb{C}^{n \times m} \rightarrow \mathbb{C}$, we have $n \cdot m$ component-wise \mathbb{R} -derivatives $\frac{\partial f}{\partial (\mathbf{Z})_{ij}}$ and $n \cdot m$ conjugate \mathbb{R} -derivatives $\frac{\partial f}{\partial (\mathbf{Z}^*)_{ij}}$. The question naturally arises how to order these $n \cdot m$ complex derivatives. Obviously this can be done in many ways, corresponding to the pattern chosen to arrange the $n \cdot m$ elements. It is worthwhile noticing that, even though they all contain the same $n \cdot m$ derivatives, not all definitions have the same properties; for instance for some of them a useful chain rule does not exist. Next, we introduce two definitions, both useful for our derivations and widely used in the literature [40]; in the former definition, the $n \cdot m$ (conjugate) \mathbb{R} -derivatives are displayed in the same order as $(\mathbf{Z})_{ij}$ and $(\mathbf{Z}^*)_{ij}$ appear in \mathbf{Z} and \mathbf{Z}^* , whereas in the latter we arrange all the elements in a row vector. Given $f : \mathbb{C}^{n \times m} \rightarrow \mathbb{C}$, the (matrix) gradient and co(nj)ugate-gradient of f at $\mathbf{Z}_0 \in \mathbb{C}^{n \times m}$ are defined as

$$\begin{aligned} \nabla_{\mathbf{Z}} f(\mathbf{Z}_0) &\triangleq \left. \frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}} \right|_{\mathbf{Z}=\mathbf{Z}_0} \quad \text{with} \quad \left[\frac{\partial f}{\partial \mathbf{Z}} \right]_{ij} = \frac{\partial f}{\partial (\mathbf{Z})_{ij}}, \quad \forall i = 1, \dots, n \quad \text{and} \quad j = 1, \dots, m \\ \nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0) &\triangleq \left. \frac{\partial f(\mathbf{Z})}{\partial \mathbf{Z}^*} \right|_{\mathbf{Z}=\mathbf{Z}_0} \quad \text{with} \quad \left[\frac{\partial f}{\partial \mathbf{Z}^*} \right]_{ij} = \frac{\partial f}{\partial (\mathbf{Z}^*)_{ij}}, \quad \forall i = 1, \dots, n \quad \text{and} \quad j = 1, \dots, m, \end{aligned} \quad (38)$$

where $\frac{\partial f}{\partial (\mathbf{Z})_{ij}}$ and $\frac{\partial f}{\partial (\mathbf{Z}^*)_{ij}}$ are the \mathbb{R} -derivative and conjugate \mathbb{R} -derivative of the complex-valued function f w.r.t. $(\mathbf{Z})_{ij}$ and $(\mathbf{Z}^*)_{ij}$, respectively. Note that $\nabla_{\mathbf{Z}} f(\mathbf{Z}_0)$ and $\nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0)$ are matrices having the same size of \mathbf{Z} . Alternatively, one can arrange the elements $\frac{\partial f}{\partial (\mathbf{Z})_{ij}}$ and $\frac{\partial f}{\partial (\mathbf{Z}^*)_{ij}}$ in a row vector, and define $D_{\mathbf{Z}} f(\mathbf{Z})$ and $D_{\mathbf{Z}^*} f(\mathbf{Z})$ at $\mathbf{Z}_0 \in \mathbb{C}^{n \times m}$ as

$$D_{\mathbf{Z}} f(\mathbf{Z}_0) \triangleq \left. \frac{\partial f(\mathbf{Z})}{\partial \text{vec}(\mathbf{Z})^T} \right|_{\mathbf{Z}=\mathbf{Z}_0} = \text{vec}(\nabla_{\mathbf{Z}} f(\mathbf{Z}_0))^T \quad \text{and} \quad D_{\mathbf{Z}^*} f(\mathbf{Z}_0) \triangleq \left. \frac{\partial f(\mathbf{Z})}{\partial \text{vec}(\mathbf{Z}^*)^T} \right|_{\mathbf{Z}=\mathbf{Z}_0} = \text{vec}(\nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0))^T, \quad (39)$$

where $\text{vec}(\mathbf{A})^T$ stands for $(\text{vec}(\mathbf{A}))^T$. For (complex-valued) *matrix* functions of complex matrices, $\mathbf{F}^{\mathbb{C}} : \mathbb{C}^{n \times m} \rightarrow \mathbb{C}^{p \times q}$, we arrange the $pq \cdot nm$ (conjugate) \mathbb{R} -derivatives in the following $pq \times nm$ matrices

$$\begin{aligned} D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}_0) &\triangleq \left. \frac{\partial \text{vec}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z}))}{\partial \text{vec}(\mathbf{Z})^T} \right|_{\mathbf{Z}=\mathbf{Z}_0}, \quad \text{with} \quad \left[\frac{\partial \text{vec}(\mathbf{F}^{\mathbb{C}})}{\partial \text{vec}(\mathbf{Z})^T} \right]_{ij} = \frac{\partial [\text{vec}(\mathbf{F}^{\mathbb{C}})]_i}{\partial [\text{vec}(\mathbf{Z})]_j}, \quad \forall i = 1, \dots, pq \quad \text{and} \quad j = 1, \dots, nm, \\ D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}_0) &\triangleq \left. \frac{\partial \text{vec}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z}))}{\partial \text{vec}(\mathbf{Z}^*)^T} \right|_{\mathbf{Z}=\mathbf{Z}_0}, \quad \text{with} \quad \left[\frac{\partial \text{vec}(\mathbf{F}^{\mathbb{C}})}{\partial \text{vec}(\mathbf{Z}^*)^T} \right]_{ij} = \frac{\partial [\text{vec}(\mathbf{F}^{\mathbb{C}})]_i}{\partial [\text{vec}(\mathbf{Z}^*)]_j}, \quad \forall i = 1, \dots, pq \quad \text{and} \quad j = 1, \dots, nm. \end{aligned} \quad (40)$$

The $pq \times nm$ matrices $D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}$ and $D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}$ are called Jacobian and conjugate Jacobian of $\mathbf{F}^{\mathbb{C}}$. Note that when $\mathbf{F}^{\mathbb{C}}$ is a *scalar* function of \mathbf{Z} , i.e., $\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) = f(\mathbf{Z})$ with $f : \mathbb{C}^{n \times m} \rightarrow \mathbb{C}$, definitions (40) reduce to (39). Practical

rules to compute \mathbb{R} -derivatives and conjugate \mathbb{R} -derivatives introduced above for specific functions of vector and matrix variables can be found in [40].

6.2 The minimum principle

Let $\mathbb{C}^{n \times m}$ be the space of complex $n \times m$ matrices, and let $\mathcal{K} \subseteq \mathbb{C}^{n \times m}$ be a closed and convex set. We consider the optimization problem

$$\begin{aligned} & \underset{\mathbf{Z}}{\text{minimize}} && f(\mathbf{Z}) \\ & \text{subject to} && \mathbf{Z} \in \mathcal{K}, \end{aligned} \tag{41}$$

where $f : \mathcal{K} \rightarrow \mathbb{R}$ is a real-valued convex and continuously \mathbb{R} -differentiable function on \mathcal{K} . At the basis of the minimum principle there is the first-order Taylor expansion of f at $\mathbf{Z}_0 \in \mathcal{K}$ as proved in Appendix E:⁷

$$\begin{aligned} f(\mathbf{Z}_0 + \Delta\mathbf{Z}) - f(\mathbf{Z}_0) &\simeq 2 \operatorname{Re} \left(\operatorname{tr} \left((\nabla_{\mathbf{Z}} f(\mathbf{Z}_0))^T \Delta\mathbf{Z} \right) \right) \\ &= 2 \operatorname{Re} \left(\operatorname{tr} (\Delta\mathbf{Z}^H (\nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0))) \right) \\ &\triangleq 2 \langle \Delta\mathbf{Z}, \nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0) \rangle \end{aligned} \tag{42}$$

where we used $(\nabla_{\mathbf{Z}} f)^* = \nabla_{\mathbf{Z}^*} f$ since f is real [see (37)], and in the last equality we introduced the inner product $\langle \bullet, \bullet \rangle : \mathbb{C}^{n \times n} \times \mathbb{C}^{n \times n} \rightarrow \mathbb{R}$, defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle \triangleq \operatorname{Re} \left(\operatorname{tr} (\mathbf{A}^H \mathbf{B}) \right). \tag{43}$$

Note that the norm induced by the inner product $\langle \bullet, \bullet \rangle$ is the Frobenius norm, i.e., $\langle \mathbf{A}, \mathbf{A} \rangle = \operatorname{Tr}(\mathbf{A}^H \mathbf{A}) = \|\mathbf{A}\|_F^2$. Using (42) we can now introduce the minimum principle as given next.

Lemma 24 *Given the convex optimization problem (41) in the setting above, $\mathbf{X} \in \mathcal{K}$ is an optimal solution of (41) if and only if \mathbf{X} satisfies $\langle \mathbf{Z} - \mathbf{X}, \nabla_{\mathbf{Z}^*} f(\mathbf{X}) \rangle \geq 0$ for all $\mathbf{Z} \in \mathcal{K}$.*

Proof. See Appendix E. ■

It is interesting to observe that if the optimal solution \mathbf{X} is in the interior of \mathcal{K} [e.g., the optimization problem (41) is unconstrained, implying $\mathcal{K} = \mathbb{C}^{n \times m}$], then the above optimality conditions reduce to $\nabla_{\mathbf{Z}^*} f(\mathbf{X}) = \mathbf{0}$, or equivalently $\nabla_{\mathbf{Z}} f(\mathbf{X}) = \mathbf{0}$, which are the well-established complex gradient-vanishing conditions obtained in [40] for the *unconstrained* minimization of real functions of complex variables. Here, we generalize that result to the case of *constrained* minimization of real-valued functions of complex matrices. In fact, Lemma 24 provides a platform for studying constrained optimization problems within the complex variables domain, taking advantage of the multivariate $\mathbb{C}\mathbb{R}$ -calculus (and in particular complex matrix derivative rules [40]) as well as the structure of the problem under consideration. We conclude this section with an example of application of the minimum principle, which is instrumental for the analysis in Sec. 7.

⁷The proposed Taylor expansion can be rewritten as the one reported in [40, Ch. 5], for which however there is no proof.

Example 25 (An application of the minimum principle) Consider the following single-user rate maximization problem

$$\begin{aligned} & \underset{\mathbf{Z}}{\text{maximize}} && f(\mathbf{Z}) \triangleq \log \det (\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H) \\ & \text{subject to} && \mathbf{Z} \in \mathcal{K}, \end{aligned} \quad (44)$$

where $\mathbf{R}_n \in \mathbb{C}^{m \times m}$ is a positive definite matrix, $\mathbf{H} \in \mathbb{C}^{m \times n}$, and \mathcal{K} is any convex and compact subset of the $n \times n$ complex positive semidefinite matrices (assumed to be nonempty). Note that $f(\mathbf{Z})$ is a concave (real-valued) function on the feasible set \mathcal{K} ; however, f is not real if defined on $\mathbb{C}^{n \times n}$. Since we are interested in minimizing f over \mathcal{K} , in order to apply the minimum principle, one approach we can follow is to consider without loss of generality the modified function $\tilde{f} : \tilde{\mathcal{K}} \rightarrow \mathbb{R}$, defined as $\tilde{f}(\mathbf{Z}) \triangleq 2 \operatorname{Re}(f(\mathbf{Z}))$, where $\tilde{\mathcal{K}} \subseteq \mathbb{C}^{n \times n}$ is any open set over which $f(\mathbf{Z})$ is well defined (it is sufficient that $\det(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H) \neq 0$); indeed, \tilde{f} coincides with f over \mathcal{K} , but it is real everywhere (in its domain). Moreover, \tilde{f} is \mathbb{R} -differentiable on $\tilde{\mathcal{K}}$.⁸ The conjugate (matrix) \mathbb{R} -derivative of \tilde{f} at $\mathbf{Z}_0 \in \tilde{\mathcal{K}}$ is (see Appendix F)

$$\nabla_{\mathbf{Z}^*} \tilde{f}(\mathbf{Z}_0) = \mathbf{H}^H (\mathbf{R}_n + \mathbf{H}\mathbf{Z}_0^H \mathbf{H}^H)^{-1} \mathbf{H}. \quad (45)$$

Introducing the complex-value matrix function $\mathbf{G} : \mathcal{K} \rightarrow \mathbb{C}^{n \times n}$, defined as $\mathbf{G} = \mathbf{G}(\mathbf{Z}) \triangleq -\nabla_{\mathbf{Z}^*} \tilde{f}(\mathbf{Z}) = -\mathbf{H}^H (\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1} \mathbf{H}$ (note that $\mathbf{Z} \in \mathcal{K}$ and thus $\mathbf{Z} = \mathbf{Z}^H$), and invoking Lemma 24, the optimization problem (44) is then equivalent to the minimum principle in the complex domain: find a $\mathbf{Z} \in \mathcal{K}$ such that $\langle \mathbf{Y} - \mathbf{Z}, \mathbf{G}(\mathbf{Z}) \rangle \geq 0$ for all $\mathbf{Y} \in \mathcal{K}$. \square

6.3 The VI problem in the complex domain

With the developments of the previous section on hand, we can now introduce the definition of the VI problem in the domain of complex matrices, termed the *complex VI problem*. Similarly to what we have done for the real case (cf. Sec. 3), we can think of the VI problem as the generalization of the minimum principle (cf. Lemma 24), where the co-gradient $\nabla_{\mathbf{Z}^*} f$ is replaced with a complex-valued matrix mapping. The formal definition is given next.

Definition 26 Given a convex and closed set $\mathcal{K} \subseteq \mathbb{C}^{n \times m}$ and a complex-valued matrix function $\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) : \mathcal{K} \ni \mathbf{Z} \rightarrow \mathbb{C}^{n \times m}$, the complex VI problem, denoted by $\text{VI}(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$, consists in finding a point $\mathbf{Z} \in \mathcal{K}$ such that $\langle \mathbf{Y} - \mathbf{Z}, \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \rangle \geq 0$ for all $\mathbf{Y} \in \mathcal{K}$. The solution set of the $\text{VI}(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$ is denoted by $\text{SOL}(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$.

When the feasible set \mathcal{K} has a Cartesian structure, i.e., $\mathcal{K} \triangleq \prod_{i=1}^I \mathcal{K}_i$ with each $\mathcal{K}_i \subseteq \mathbb{C}^{n_i \times m_i}$, we write $\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \triangleq (\mathbf{F}_i^{\mathbb{C}}(\mathbf{Z}))_{i=1}^I$ and $\mathbf{Z} \triangleq (\mathbf{Z}_i)_{i=1}^I$, with $\mathbf{F}_i^{\mathbb{C}}(\mathbf{Z}) : \mathcal{K} \rightarrow \mathbb{C}^{n_i \times m_i}$ and $\mathbf{Z}_i \in \mathbb{C}^{n_i \times m_i}$. In such a

⁸The introduction of the auxiliary function \tilde{f} might appear an unnecessary complication, which needs clarification. The original function f is defined over a (sub)set of positive semidefinite matrices. The theory of matrix derivatives introduced in this paper cannot be applied directly to functions of matrices having a structure. The function \tilde{f} allows us to overcome this issue, because it is defined over an open set of *unpatterned* matrices (over which we can use the developed matrix differentiation theory) while being equal to f over the set of interest. A different approach would be working directly with the original function f and using the so-called complex (patterned) generalized derivatives [40]. However, state-of-the-art results show that it is *not* possible to find generalized derivatives w.r.t. an *arbitrary* set of complex-valued patterned matrices, which strongly limits the applicability of this methodology in practice. Using instead the approach explained in this example, the proposed matrix differentiation machinery can be applied also to optimization problems involving patterned matrices.

case, with a slight abuse of notation, we will still use for the partitioned VI $(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$ the compact notation $\langle \mathbf{Y} - \mathbf{Z}, \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \rangle \geq 0$, by meaning $\sum_{i=1}^I \langle \mathbf{Y}_i - \mathbf{Z}_i, \mathbf{F}_i^{\mathbb{C}}(\mathbf{Z}) \rangle \geq 0$. Moreover, the definitions of $D_{\mathbf{Z}}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ and $D_{\mathbf{Z}^*}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ as given in (40) depend in principle on the ordering according to which the components of $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ and \mathbf{Z} are grouped in the vec operator. For our purposes, the following ordering is the most convenient, which is tacitly assumed throughout the paper: $\text{vec}((\mathbf{F}_i^{\mathbb{C}}(\mathbf{Z}))_{i=1}^I) \triangleq [\text{vec}(\mathbf{F}_1^{\mathbb{C}}(\mathbf{Z}))^T, \dots, \text{vec}(\mathbf{F}_I^{\mathbb{C}}(\mathbf{Z}))^T]^T$ and $\text{vec}((\mathbf{Z}_i)_{i=1}^I) \triangleq [\text{vec}(\mathbf{Z}_1)^T, \dots, \text{vec}(\mathbf{Z}_I)^T]^T$.

6.4 Monotonicity and P properties of VI $(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$

We can now readily extend the definitions of monotonicity and P property as introduced in Proposition 16 to the complex-value matrix mapping $\mathbf{F}^{\mathbb{C}}$; the aforementioned definitions are in fact formally the same, with the only difference that the scalar product and the Euclidean norm are replaced with the inner product $\langle \bullet, \bullet \rangle$ defined in (43) and the Frobenius norm, respectively. The non-trivial task is instead to provide conditions easy to check that guarantee these properties. These conditions are indeed instrumental to study convergence of algorithms for complex NEPs. The interesting result we prove next is that we can obtain necessary and sufficient conditions for a continuously (\mathbb{R} -)differentiable $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ to be a monotone function or (sufficient conditions to be) a P function on \mathcal{K} that are formally equivalent to those obtained for real-valued vector functions $\mathbf{F}(\mathbf{x})$ [cf. (20) and Proposition 5], provided that we introduce a novel definition of Jacobian matrix suitable for complex-valued functions of complex variables; such a Jacobian will contain both \mathbb{R} -derivatives and conjugate \mathbb{R} -derivatives of $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$.

Given the complex VI $(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$, suppose that $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ is a continuously (\mathbb{R} -)differentiable matrix function on \mathcal{K} . Then, the $nm \times nm$ Jacobian matrices $D_{\mathbf{Z}}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ and $D_{\mathbf{Z}^*}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ in (40) are well-defined at $\mathbf{Z} \in \mathcal{K}$. Let us introduce the $2nm \times 2nm$ matrix $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$, defined as

$$\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \triangleq \frac{1}{2} \begin{bmatrix} D_{\mathbf{Z}}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) & D_{\mathbf{Z}^*}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \\ D_{\mathbf{Z}}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*) & D_{\mathbf{Z}^*}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*) \end{bmatrix}, \quad (46)$$

which we call “augmented Jacobian” for obvious reasons. For notational simplicity, in the sequel we will write $D_{\mathbf{Z}}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*$ and $D_{\mathbf{Z}^*}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*$ for $D_{\mathbf{Z}}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*)$ and $D_{\mathbf{Z}^*}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*)$, respectively. Note that the following relationships hold between the blocks of $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$: $(D_{\mathbf{Z}}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}))^* = D_{\mathbf{Z}^*}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*$ and $(D_{\mathbf{Z}^*}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}))^* = D_{\mathbf{Z}}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})^*$. Finally, under the assumption that $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ and \mathcal{K} have a partitioned structure and $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ has bounded (\mathbb{R})-derivatives on \mathcal{K} , let us introduce the “condensed” $I \times I$ matrix $\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}$ given by

$$[\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}]_{ij} \triangleq \begin{cases} \kappa_i^{\min}, & \text{if } i = j, \\ -\xi_{ij}^{\max}, & \text{otherwise,} \end{cases} \quad (47)$$

with

$$\kappa_i^{\min} \triangleq \inf_{\mathbf{Z} \in \mathcal{K}} \lambda_{\text{least}} \left(\mathbf{A}_i^H \mathbf{J}_i \mathbf{F}_i^{\mathbb{C}}(\mathbf{Z}) \mathbf{A}_i \right) \quad \text{and} \quad \xi_{ij}^{\max} \triangleq \sup_{\mathbf{Z} \in \mathcal{K}} \left\| \mathbf{A}_i^H \mathbf{J}_j \mathbf{F}_i^{\mathbb{C}}(\mathbf{Z}) \mathbf{A}_j \right\|_F, \quad (48)$$

where $\mathbf{J}_i \mathbf{F}_i^{\mathbb{C}}(\mathbf{Z})$ and $\mathbf{J}_j \mathbf{F}_i^{\mathbb{C}}(\mathbf{Z})$ represent the augmented Jacobian matrices of $\mathbf{F}_i^{\mathbb{C}}(\mathbf{Z})$ as defined in (46), whose \mathbb{R} -derivatives are taken with respect to the matrix variables \mathbf{Z}_i and \mathbf{Z}_j (and their conjugates), respectively; $\mathbf{A}_i \in \mathbb{C}^{2n_i m_i \times 2n_i m_i}$ are nonsingular arbitrary matrices; and $\|\mathbf{A}\|_F$ denotes the Frobenius norm of \mathbf{A} . Matrices $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ and $\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}$ play for complex VIs the same role as $\mathbf{J}\mathbf{F}$ and $\mathbf{\Upsilon}_{\mathbf{F}}$ introduced in Sec. 3.1 for real VIs. Monotonicity and P properties of VI $(\mathcal{K}, \mathbf{F}^{\mathbb{C}})$ are indeed relayed on those matrices, as shown next.

Before stating the main results (Propositions 28 and 29), we need to introduce a novel relaxed definition of (uniformly) positive (semi-)definiteness for matrices in the form (46), which takes explicitly into account the special structure of those matrices. Instead of checking the sign of the quadratic form $\mathbf{y}^H \mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \mathbf{y}$ for all possible vectors $\mathbf{y} \in \mathbb{C}^{2nm}$, it turns out that we can restrict the check to structured vectors in the form $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_1^*]$ for all $\mathbf{y}_1 \in \mathbb{C}^{nm}$, which is actually the size of the vector space where \mathbf{Z} lies (see Propositions 28 and 29 below). This motivates the following definition of “augmented” (uniformly) positive (semi-)definiteness for matrices in the form of (46).

Definition 27 *The augmented Jacobian $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is said to be:*

i) augmented positive semidefinite on \mathcal{K} if for all $\mathbf{Y} \in \mathbb{C}^{n \times m}$ and $\mathbf{Z} \in \mathcal{K}$,

$$\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq 0; \quad (49)$$

ii) augmented positive definite on \mathcal{K} if for all $\mathbf{0} \neq \mathbf{Y} \in \mathbb{C}^{n \times m}$ and $\mathbf{Z} \in \mathcal{K}$, the inequality in (49) is strict;

iii) uniformly augmented positive definite on \mathcal{K} with constant $c > 0$ if for all $\mathbf{Y} \in \mathbb{C}^{n \times m}$ and $\mathbf{Z} \in \mathcal{K}$, there exists a positive constant $c > 0$ such that

$$\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq c \|\mathbf{Y}\|_F^2. \quad (50)$$

For i), ii), and iii) we will write $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \stackrel{A}{\succeq} \mathbf{0}$, $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \stackrel{A}{\succ} \mathbf{0}$, and $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) - c\mathbf{I} \stackrel{A}{\succeq} \mathbf{0}$, respectively.

Note that $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is not Hermitian; which implies that $\text{vec}(\mathbf{W})^H \mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \text{vec}(\mathbf{W})$ is generally not a real number for arbitrary $\mathbf{W} \in \mathbb{C}^{n \times 2m}$. However, because of the structure of $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ and $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])$, with $\mathbf{Y} \in \mathbb{C}^{n \times m}$, the quadratic form $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{JF}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*])$ introduced in the proposition is always real. Note also that if $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is positive semidefinite, positive definite, or uniformly positive definite on \mathcal{K} (and thus Hermitian), then it is also augmented positive semidefinite, positive definite, or uniformly positive definite, respectively; but the converse in general is not true (because $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is not Hermitian). Using this new relaxed concept of positive definiteness, we can now establish the connection with the monotonicity properties of $\mathbf{F}^{\mathbb{C}}$.

Proposition 28 *Let $\mathbf{F}^{\mathbb{C}} : \mathcal{K} \rightarrow \mathbb{C}^{n \times m}$ be (\mathbb{R}) -continuously differentiable on the convex set \mathcal{K} . Suppose that \mathcal{K} has nonempty interior. The following statements hold:*

- (a) $\mathbf{F}^{\mathbb{C}}$ is monotone on \mathcal{K} if and only if $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is augmented positive semidefinite on \mathcal{K} ;
- (b) If $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is augmented positive definite on \mathcal{K} , then $\mathbf{F}^{\mathbb{C}}$ is strictly monotone on \mathcal{K} ;
- (c) $\mathbf{F}^{\mathbb{C}}$ is strongly monotone on \mathcal{K} with constant $c_{sm} > 0$ if and only if $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ is uniformly augmented positive definite on \mathcal{K} with constant $c_{sm}/2$.

If we assume a Cartesian product structure, i.e. $\mathbf{F}^{\mathbb{C}} = (\mathbf{F}_i^{\mathbb{C}})_{i=1}^I$ and $\mathcal{K} = \prod_{i=1}^I \mathcal{K}_i$, and bounded (\mathbb{R}) -derivatives of $\mathbf{F}^{\mathbb{C}}$ on \mathcal{K} , then:

- (d) If $\mathbf{Y}_{\mathbf{F}^{\mathbb{C}}}$ is positive semidefinite/ P_0 -matrix, then $\mathbf{F}^{\mathbb{C}}$ is a monotone/ P_0 function on \mathcal{K} ;
- (e) If $\mathbf{Y}_{\mathbf{F}^{\mathbb{C}}}$ is a P -matrix, then $\mathbf{F}^{\mathbb{C}}$ is a uniformly P -function on \mathcal{K} .

Proof. See Appendix G. ■

The above proposition is the generalization of (20) and Proposition 5 to complex VIs. Note that, if the set \mathcal{K} has empty interior, necessary conditions in (a) and (c) generally do not hold, whereas sufficient conditions in (a)-(c) may be too restrictive. Since some of the optimization problems of our interest have feasible sets that fall into this class [e.g., think of the set of Hermitian matrices], it is worth extending Proposition 28 to sets with empty interior. The next result is valid for arbitrary (nonempty) convex sets.

Proposition 29 *Consider the setting of Proposition 28, but with \mathcal{K} being any nonempty convex subset of $\mathbb{C}^{n \times m}$. Let $\mathcal{S}_{\mathcal{K}}$ be the subspace that is parallel to the affine hull of \mathcal{K} .⁹ The following statements hold:*

- (a) $\mathbf{F}^{\mathbb{C}}$ is monotone on \mathcal{K} if and only if for all $\mathbf{Y} \in \mathbb{C}^{n \times m}$ such that $\mathbf{Y} \in \mathcal{S}_{\mathcal{K}}$ and $\mathbf{Z} \in \mathcal{K}$, it holds $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq 0$
- (b) If for all $\mathbf{0} \neq \mathbf{Y} \in \mathbb{C}^{n \times m}$ such that $\mathbf{Y} \in \mathcal{S}_{\mathcal{K}}$ and $\mathbf{Z} \in \mathcal{K}$, it holds $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) > 0$, then $\mathbf{F}^{\mathbb{C}}$ is strictly monotone on \mathcal{K} ;
- (c) $\mathbf{F}^{\mathbb{C}}$ is strongly monotone on \mathcal{K} with constant $c_{sm} > 0$ if and only if for all $\mathbf{Y} \in \mathbb{C}^{n \times m}$ such that $\mathbf{Y} \in \mathcal{S}_{\mathcal{K}}$ and $\mathbf{Z} \in \mathcal{K}$, it holds $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq (c_{sm}/2) \|\mathbf{Y}\|_F^2$.

If we assume a Cartesian product structure, i.e. $\mathbf{F}^{\mathbb{C}} = (\mathbf{F}_i^{\mathbb{C}})_{i=1}^I$ and $\mathcal{K} = \prod_{i=1}^I \mathcal{K}_i$, and bounded (\mathbb{R}) -derivatives of $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ on \mathcal{K} , then statements (d) and (e) of Proposition 28 hold.

Proof. See Appendix G. ■

A set \mathcal{K} of special interest for our applications (cf. Sec. 2) is the set of complex $n \times n$ positive semidefinite matrices (and thus Hermitian). This set has empty interior, implying that one needs to use Proposition 29. It is not difficult to see that the affine hull of such a \mathcal{K} is the set of Hermitian matrices, which is already a subspace. Therefore, when Proposition 29 applies to such a \mathcal{K} , the matrices \mathbf{Y} are restricted to the set of Hermitian matrices. It is worth observing that, when \mathcal{K} has nonempty interior, Proposition 29 reduces to Proposition 28; indeed, we have $\text{Aff}(\mathcal{K}) = \mathbb{C}^{n \times m}$, and thus $\mathcal{S}_{\mathcal{K}} = \mathbb{C}^{n \times m}$.

Using Proposition 28 (or Proposition 29) and building on the structure of $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ one can obtain sufficient conditions for $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ to be augmented positive (semi-)definite or uniformly positive semidefinite, similarly to what we have done in Sec. 3.1 for real valued vector functions $\mathbf{F}(\mathbf{x})$; one can then extend the solution analysis and methods developed for the VI(\mathcal{Q}, \mathbf{F}) to the complex VI($\mathcal{K}, \mathbf{F}^{\mathbb{C}}$); because of the space limitation, we leave these tasks to the reader. In Sec. 7, we will show an instance of these conditions when specialized to the MIMO games along with their physical interpretations.

We conclude this section by applying Proposition 28 (or Proposition 29) to the conjugate gradient of real-valued functions of complex variables [cf. (41)]. The result is a set of novel necessary and sufficient

⁹We recall that, given a subset \mathcal{K} of $\mathbb{C}^{n \times m}$, the affine hull of \mathcal{K} , denoted by $\text{Aff}(\mathcal{K})$, is the set of all affine combinations of elements in \mathcal{K} , that is $\text{Aff}(\mathcal{K}) \triangleq \left\{ \mathbf{Y} \in \mathbb{C}^{n \times m} : \mathbf{Y} = \sum_{i=1}^k \alpha_i \mathbf{X}_i, k > 0, \mathbf{X}_i \in \mathcal{K}, \alpha_i \in \mathbb{R}, \sum_{i=1}^k \alpha_i = 1 \right\}$.

conditions for a (continuously \mathbb{R} -differentiable) real-valued function of complex variables to be (strictly) convex or strongly convex, in terms of \mathbb{R} -derivatives. This provides an easy way to check convexity directly in the complex domain. In order to apply Propositions 28 or 29, we need the following intermediate result, which can be proved using the Taylor expansion (42) and following the same approach used to prove similar results for real-valued functions of real variables. Given a continuously \mathbb{R} -differentiable real-valued function $f : \mathbb{C}^{n \times m} \rightarrow \mathbb{R}$, f is convex, strictly convex, or strongly convex on \mathcal{K} if and only if its conjugate gradient $\nabla_{\mathbf{Z}^*} f(\mathbf{Z})$ is monotone, strictly monotone, or strongly monotone on \mathcal{K} , respectively. Using Proposition 28 (or Proposition 29), the convexity properties of $f(\mathbf{Z})$ can be then restated in terms of properties of the augmented Jacobian matrix $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ of $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$, with $\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) = \nabla_{\mathbf{Z}^*} f(\mathbf{Z})$, which we term *augmented Hessian* of $f(\mathbf{Z})$, $\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z})$, given by [cf. (46)]:

$$\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z}) \triangleq \frac{1}{2} \begin{bmatrix} D_{\mathbf{Z}}(\nabla_{\mathbf{Z}^*} f(\mathbf{Z})) & D_{\mathbf{Z}^*}(\nabla_{\mathbf{Z}^*} f(\mathbf{Z})) \\ D_{\mathbf{Z}}((\nabla_{\mathbf{Z}^*} f(\mathbf{Z}))^*) & D_{\mathbf{Z}^*}((\nabla_{\mathbf{Z}^*} f(\mathbf{Z}))^*) \end{bmatrix} \triangleq \frac{1}{2} \begin{bmatrix} \nabla_{\mathbf{Z}\mathbf{Z}^*}^2 f(\mathbf{Z}) & \nabla_{\mathbf{Z}^*\mathbf{Z}^*}^2 f(\mathbf{Z}) \\ \nabla_{\mathbf{Z}\mathbf{Z}}^2 f(\mathbf{Z}) & \nabla_{\mathbf{Z}^*\mathbf{Z}}^2 f(\mathbf{Z}) \end{bmatrix}. \quad (51)$$

Note that [cf. (40)] $\nabla_{\mathbf{Z}\mathbf{Z}^*}^2 f(\mathbf{Z}) = (\nabla_{\mathbf{Z}^*\mathbf{Z}}^2 f(\mathbf{Z}))^*$ and $\nabla_{\mathbf{Z}^*\mathbf{Z}^*}^2 f(\mathbf{Z}) = (\nabla_{\mathbf{Z}\mathbf{Z}}^2 f(\mathbf{Z}))^*$. It follows from Proposition 28 applied to $\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) = \nabla_{\mathbf{Z}^*} f(\mathbf{Z})$ that $\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z})$ plays the role of the classical Hessian matrix of f : let $\mathcal{K} \subseteq \mathbb{C}^{n \times m}$ be any convex set with nonempty interior, then

$$\begin{aligned} f(\mathbf{Z}) \text{ is convex on } \mathcal{K} &\Leftrightarrow \mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z}) \stackrel{A}{\succeq} \mathbf{0}, \forall \mathbf{Z} \in \mathcal{K}; \\ f(\mathbf{Z}) \text{ is strictly convex } \mathcal{K} &\Leftarrow \mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z}) \stackrel{A}{\succ} \mathbf{0}, \forall \mathbf{Z} \in \mathcal{K}; \\ f(\mathbf{Z}) \text{ is strongly convex on } \mathcal{K} &\Leftrightarrow \mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z}) - c_{\text{sm}} \mathbf{I} \stackrel{A}{\succeq} \mathbf{0}, \forall \mathbf{Z} \in \mathcal{K} \text{ and some } c_{\text{sm}} > 0. \end{aligned} \quad (52)$$

If the set \mathcal{K} has empty interior, conditions (52) are replaced by those in Proposition 29 applied to $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Z}) = \mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z})$; we leave this easy task to the reader. Note that our conditions in (52) (and Proposition 29) generalize those obtained in [58, Prop. 1.2.6 and Exercise 1.8] for real-valued functions of real variables, and sufficient conditions reported in [40, Ch. 5.3.1] for the convexity (only) of real-valued functions of complex variables defined over $\mathbb{C}^{n \times m}$ (i.e., $\mathcal{K} = \mathbb{C}^{n \times m}$).

Example 25 Revisited. Going back to the optimization problem (44), we can recover the well-known concavity property of $f(\mathbf{Z})$ on the compact and convex set \mathcal{K} by a direct application of Proposition 29. The expression of the augmented Hessian of $f(\mathbf{Z})$ will be also used in Sec. 7.2 for studying MIMO games.

Let $\tilde{\mathcal{K}}$ be any open set containing \mathcal{K} over which $f(\mathbf{Z})$ is well defined, and let $\tilde{f} : \tilde{\mathcal{K}} \rightarrow \mathbb{R}$ be $\tilde{f}(\mathbf{Z}) \triangleq 2 \operatorname{Re}(f(\mathbf{Z}))$. Since $\tilde{f} = f$ on \mathcal{K} , concavity of f on \mathcal{K} follows from that of \tilde{f} on \mathcal{K} . Since \mathcal{K} has empty interior, one needs to use Proposition 29. Observing that, for the specific set \mathcal{K} under consideration, the set $S_{\mathcal{K}}$ in Proposition 29 is $S_{\mathcal{K}} = \{\mathbf{X} \in \mathbb{C}^{n \times n} : \mathbf{X} = \mathbf{X}^H\}$, it is sufficient to show that

$$-\begin{bmatrix} \operatorname{vec}(\mathbf{Y}) \\ \operatorname{vec}(\mathbf{Y}^*) \end{bmatrix}^H \mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} \tilde{f}(\mathbf{Z}) \begin{bmatrix} \operatorname{vec}(\mathbf{Y}) \\ \operatorname{vec}(\mathbf{Y}^*) \end{bmatrix} \geq 0, \quad \forall \mathbf{Z} \in \mathcal{K} \text{ and } \forall \mathbf{Y} = \mathbf{Y}^H, \quad (53)$$

where $\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} \tilde{f}(\mathbf{Z})$ is the augmented Hessian of $\tilde{f}(\mathbf{Z})$. In Appendix F, we show that

$$\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} \tilde{f}(\mathbf{Z}) = -\frac{1}{2} \begin{bmatrix} \mathbf{0} & [\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})] \mathbf{K}_{n^2 n^2} \\ [\mathbf{G}(\mathbf{Z})^H \otimes \mathbf{G}^*(\mathbf{Z})] \mathbf{K}_{n^2 n^2} & \mathbf{0} \end{bmatrix} \quad (54)$$

where $\mathbf{G}(\mathbf{Z}) \triangleq \mathbf{H}^H (\mathbf{R}_n + \mathbf{H}\mathbf{Z}^H\mathbf{H}^H)^{-1} \mathbf{H}$ and $\mathbf{K}_{n^2n^2}$ is an $n^2 \times n^2$ permutation matrix such that $\text{vec}(\mathbf{Z}^T) = \mathbf{K}_{n^2n^2} \text{vec}(\mathbf{Z})$ (also termed commutation matrix [40, Def. 2.9]). Using (54), condition (53) becomes

$$\begin{aligned}
0 &\leq \begin{bmatrix} \text{vec}(\mathbf{Y}) \\ \text{vec}(\mathbf{Y}^*) \end{bmatrix}^H \begin{bmatrix} \mathbf{0} & \mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z}) \\ \mathbf{G}(\mathbf{Z})^H \otimes \mathbf{G}(\mathbf{Z})^* & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{K}_{n^2n^2} \text{vec}(\mathbf{Y}) \\ \mathbf{K}_{n^2n^2} \text{vec}(\mathbf{Y}^*) \end{bmatrix} \\
&= \begin{bmatrix} \text{vec}(\mathbf{Y}) \\ \text{vec}(\mathbf{Y}^*) \end{bmatrix}^H \begin{bmatrix} \mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z}) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}(\mathbf{Z})^H \otimes \mathbf{G}(\mathbf{Z})^* \end{bmatrix} \begin{bmatrix} \text{vec}(\mathbf{Y}^H) \\ \text{vec}(\mathbf{Y}^T) \end{bmatrix} \\
&= \begin{bmatrix} \text{vec}(\mathbf{Y}) \\ \text{vec}(\mathbf{Y}^*) \end{bmatrix}^H \begin{bmatrix} \mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z}) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}(\mathbf{Z})^H \otimes \mathbf{G}(\mathbf{Z})^* \end{bmatrix} \begin{bmatrix} \text{vec}(\mathbf{Y}) \\ \text{vec}(\mathbf{Y}^*) \end{bmatrix}, \quad \forall \mathbf{Z} \in \mathcal{K} \text{ and } \forall \mathbf{Y} = \mathbf{Y}^H, \quad (55)
\end{aligned}$$

where in the second equality we used the property $\mathbf{K}_{n^2n^2} \text{vec}(\mathbf{Z}) = \text{vec}(\mathbf{Z}^T)$, whereas in the last equality the condition $\mathbf{Y} = \mathbf{Y}^H$. It turns out that (55) is satisfied if $\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})$ is positive semidefinite for all $\mathbf{Z} \in \mathcal{K}$. Since $\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})$ is Hermitian on \mathcal{K} , it is sufficient to check that the minimum eigenvalue of $\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})$, denoted by $\lambda_{\min}(\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z}))$, is nonnegative for all \mathbf{Z} on \mathcal{K} . The result follows from $\lambda_{\min}(\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})) = \lambda_{\min}(\mathbf{G}(\mathbf{Z})^T) \cdot \lambda_{\min}(\mathbf{G}(\mathbf{Z})) = \lambda_{\min}(\mathbf{G}(\mathbf{Z}))^2 \geq 0$ for all $\mathbf{Z} \in \mathcal{K}$, which proves concavity of $\tilde{f}(\mathbf{Z})$ on \mathcal{K} and thus of $f(\mathbf{Z})$ on \mathcal{K} .

It is worth observing that while $-\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*} f(\mathbf{Z})$ satisfies (53) [Proposition 29(a)], it is not positive (semi-) definite, showing that the latter condition may be too restrictive for checking the convexity of a (real-valued) function of complex variables. This strengthens the importance of the proposed new concept of augmented positive (semi-)definiteness (Definition 27) and the role of Propositions 28 and 29. \square

6.5 NEPs in the complex domain

We can now establish the formal connection between complex NEPs and complex VIs. Let $\mathcal{G}_{\mathbb{C}} \triangleq \langle \mathcal{K}, \mathbf{f} \rangle$ be a complex NEP where each player controls a complex matrix $\mathbf{Z}_i \in \mathbb{C}^{n_i \times m_i}$ that must belong to the player's feasible set $\mathcal{K}_i \subseteq \mathbb{C}^{n_i \times m_i}$; the cost function of each player is denoted by $f_i : \mathcal{K} \rightarrow \mathbb{R}$; and the joint strategy set of the game is $\mathcal{K} = \prod_i \mathcal{K}_i$. We also write $\mathbf{Z} \triangleq (\mathbf{Z}_i)_{i=1}^Q$, $\mathbf{Z}_{-i} \triangleq (\mathbf{Z}_1, \dots, \mathbf{Z}_{i-1}, \mathbf{Z}_{i+1}, \dots, \mathbf{Z}_Q)$, and $\mathcal{K}_{-i} \triangleq \prod_{j \neq i} \mathcal{K}_j$. The NEP problem $\mathcal{G}_{\mathbb{C}}$ consists then, for each player $i = 1, \dots, Q$, in solving the following convex optimization problem: given $\mathbf{Z}_{-i} \in \mathcal{K}_{-i}$,

$$\begin{aligned}
&\underset{\mathbf{Z}_i}{\text{minimize}} && f_i(\mathbf{Z}_i, \mathbf{Z}_{-i}) \\
&\text{subject to} && \mathbf{Z}_i \in \mathcal{K}_i.
\end{aligned} \quad (56)$$

Building on Lemma 24, it is not difficult to prove the following.

Proposition 30 *Given the complex NEP $\mathcal{G}_{\mathbb{C}} \triangleq \langle \mathcal{K}, \mathbf{f} \rangle$, suppose that for each player i the following hold:*

- i) *the (nonempty) strategy set \mathcal{K}_i is closed and convex;*
- ii) *the payoff function $f_i(\mathbf{Z}_i, \mathbf{Z}_{-i})$ is convex and continuously \mathbb{R} -differentiable in \mathbf{Z}_i for every fixed \mathbf{Z}_{-i} .*

Then, the complex NEP $\mathcal{G}_{\mathbb{C}}$ is equivalent to the complex VI($\mathcal{K}, \mathbf{G}^{\mathbb{C}}$), where $\mathbf{G}^{\mathbb{C}}(\mathbf{Z}) \triangleq (\nabla_{\mathbf{Z}_i^} f_i(\mathbf{Z}))_{i=1}^Q$.*

Exploring the above connection between $\mathcal{G}_{\mathbb{C}}$ and the VI($\mathcal{K}, \mathbf{G}^{\mathbb{C}}$) as well as the monotonicity/P properties of VI($\mathcal{K}, \mathbf{G}^{\mathbb{C}}$) (cf. Proposition 28 and Proposition 29), one can then extend the solution analysis and algorithms

developed for the real NEP \mathcal{G} in Sec. 3, 4 and 5 to the complex NEP $\mathcal{G}_{\mathbb{C}}$ $[\text{VI}(\mathcal{K}, \mathbf{G}^{\mathbb{C}})]$; because of space limitations, we leave this task to the reader. In the next section we will show how to explore the connection between $\mathcal{G}_{\mathbb{C}}$ and the $\text{VI}(\mathcal{K}, \mathbf{G}^{\mathbb{C}})$ along with the developed theory of complex VIs to study and solve the MIMO games introduced in Sec. 2.2.

7 Noncooperative Games Over Interference Channels Revisited

In this section, we focus on the application of the general theory developed in the previous sections to some concrete examples of practical interest. In particular, we show how the real/complex NEPs introduced in Sec. 2 can be naturally casted in the proposed framework and thus efficiently solved. The main result in the SISO case is a novel iterative water-filling like algorithm where the users can choose the degree of desired cooperation via local pricing, converging to solutions having different performance/signaling trade-off; we also prove that the best-response of each player has a multi-level water-filling-like expression and provide an efficient algorithm for its computation. We then extend our analysis to MIMO games and obtain similar results. Numerical experiments show the superiority of our novel algorithms with respect to plain noncooperative solutions as well as very good performance with respect to centralized schemes, in favor of very limited signaling among the users.

7.1 The SISO case

We study here the game $\mathcal{G}_{\text{siso}} = \langle \mathcal{P}^{\text{siso}}, (r_i)_{i=1}^I \rangle$ introduced in (7). The VI function associated with $\mathcal{G}_{\text{siso}}$ is $\mathbf{G}(\mathbf{p}) \triangleq (\mathbf{G}_i(\mathbf{p}))_{i=1}^I : \mathcal{P}^{\text{siso}} \rightarrow \mathbb{R}^{NI}$, where each $\mathbf{G}_i(\mathbf{p})$ is defined as

$$\mathbf{G}_i(\mathbf{p}) \triangleq -\nabla_{\mathbf{p}_i} r_i(\mathbf{p}) = \left(-\frac{|H_{ii}(k)|^2}{\sigma_i^2(k) + \sum_{j \neq i} |H_{ij}(k)|^2 p_j(k)} \right)_{k=1}^N. \quad (57)$$

Note that in this section, due to the nature of the problems at hand, we called the VI mapping \mathbf{G} instead of \mathbf{F} used previously, and the VI variables \mathbf{p} instead of \mathbf{x} as used previously.

According to Proposition 5, the monotonicity/P properties of $\mathbf{G}(\mathbf{p})$ are related to the matrices $\mathbf{J}\mathbf{G}_{\text{low}}$ and $\mathbf{\Upsilon}_{\mathbf{G}}$ defined in (21) and (22). We recall that the matrices \mathbf{B} and \mathbf{C}_i 's in the definition of $\mathbf{J}\mathbf{G}_{\text{low}}$ and $\mathbf{\Upsilon}_{\mathbf{G}}$ represent a degree of freedom that one can use; in this case it is convenient to make the following choices. Let us rearrange the components of \mathbf{p} by subcarriers, meaning that the vector $\mathbf{p} = (\mathbf{p}_i)_{i=1}^I$ is permuted into $\bar{\mathbf{p}} = (\bar{\mathbf{p}}(k))_{k=1}^N$, with $\bar{\mathbf{p}}(k) = (p_i(k))_{i=1}^I$; it is not difficult to see that $\bar{\mathbf{p}}$ can be written as $\bar{\mathbf{p}} = \mathbf{P}\mathbf{p}$, where \mathbf{P} is a permutation matrix defined as

$$[\mathbf{P}]_{ij} \triangleq \begin{cases} 1, & \text{if } j = [(i \bmod I) - 1]N + \lceil i/I \rceil \bmod (I \cdot N) \\ 0, & \text{otherwise.} \end{cases}$$

Using this new order of the variables, matrix $\mathbf{J}\mathbf{G}$ becomes $\mathbf{P}^T \mathbf{J}\mathbf{G}\mathbf{P}$; $\mathbf{J}\mathbf{G}_{\text{low}}$ is then obtained from $\mathbf{P}^T \mathbf{J}\mathbf{G}\mathbf{P}$ according to (21), where $\mathbf{B} \triangleq \text{Diag} \left\{ (\mathbf{B}(k))_{k=1}^N \right\}$ is a block diagonal matrix, with each block $\mathbf{B}(k) \in \mathbb{R}^{I \times I}$ being a positive diagonal matrix with the i -th entry equal to $[\mathbf{B}(k)]_{ii} \triangleq \sigma_i^2(k)/|H_{ii}(k)|^2 + \sum_j (|H_{ij}(k)|^2/|H_{ii}(k)|^2) p_j^{\max}(k)$. Matrix $\mathbf{\Upsilon}_{\mathbf{G}}$ comes directly from the original $\mathbf{J}\mathbf{G}$ by choosing each $\mathbf{C}_i \in \mathbb{R}^{N \times N}$ as a diagonal matrix, defined

as $\mathbf{C}_i \triangleq \text{Diag} \left\{ \left(\left(\sigma_i^2(k) + \sum_j |H_{ij}(k)|^2 p_j^{\max}(k) \right) / |H_{ii}(k)|^2 \right)_{k=1}^N \right\}$. The explicit expressions of $\mathbf{J}\mathbf{G}_{\text{low}}$ and $\mathbf{Y}_{\mathbf{G}}$ are the following: $\mathbf{J}\mathbf{G}_{\text{low}} \triangleq \text{Diag} \left\{ (\mathbf{J}\mathbf{G}_{\text{low}}(k))_{k=1}^N \right\} \in \mathbb{R}^{N \times N \times I}$ is a block diagonal matrix, whose k -th diagonal block $\mathbf{J}\mathbf{G}_{\text{low}}(k) \in \mathbb{R}^{I \times I}$ is

$$[\mathbf{J}\mathbf{G}_{\text{low}}(k)]_{ij} \triangleq \begin{cases} 1, & \text{if } i = j \\ -\frac{|H_{ij}(k)|^2}{|H_{jj}(k)|^2} \cdot \text{innr}_{ij}(k), & \text{if } i \neq j, \end{cases} \quad k = 1, \dots, N, \quad (58)$$

and $\mathbf{Y}_{\mathbf{G}} \in \mathbb{R}^{I \times I}$ is given by

$$[\mathbf{Y}_{\mathbf{G}}]_{ij} \triangleq \begin{cases} 1 & \text{if } i = j \\ -\max_{1 \leq k \leq N} \left\{ \frac{|H_{ij}(k)|^2}{|H_{jj}(k)|^2} \cdot \text{innr}_{ij}(k) \right\} & \text{if } i \neq j, \end{cases} \quad (59)$$

with

$$\text{innr}_{ij}(k) \triangleq \frac{\sigma_j^2(k) + \sum_r |H_{jr}(k)|^2 p_r^{\max}(k)}{\sigma_i^2(k)}. \quad (60)$$

Using the above definitions along with Proposition 5, Theorem 8, and Corollary 10, the main properties of \mathcal{G} are then given in the following proposition.

Proposition 31 *Given the real NEP $\mathcal{G}_{\text{siso}} = \langle \mathcal{P}^{\text{siso}}, (r_i)_{i=1}^I \rangle$, the following hold.*

- (a) $\mathcal{G}_{\text{siso}}$ is equivalent to the VI($\mathcal{P}^{\text{siso}}, \mathbf{G}$), which has a nonempty and compact solution set;
- (b) Suppose that $\mathbf{J}\mathbf{G}_{\text{low}}$ is positive semidefinite (positive definite). Then $\mathbf{G}(\mathbf{p})$ is monotone (strongly monotone) on $\mathcal{P}^{\text{siso}}$; therefore $\mathcal{G}_{\text{siso}}$ is a monotone NEP;
- (c) Suppose that $\mathbf{Y}_{\mathbf{G}}$ is a P-matrix (positive definite matrix). Then $\mathbf{G}(\mathbf{p})$ is a uniformly P-function (strongly monotone function) on $\mathcal{P}^{\text{siso}}$; therefore $\mathcal{G}_{\text{siso}}$ is a $P_{\mathbf{Y}}$ NEP and has a unique NE.

Using Proposition 10, we obtain the following corollary providing sufficient conditions for the matrices in Proposition 31 to be P or positive definite.

Corollary 32 *The matrix $\mathbf{Y}_{\mathbf{G}}$ in (59) is a P-matrix (or a positive definite matrix) if one (or both) of the following two sets of conditions are satisfied: for some $\mathbf{w} = (w_i)_{i=1}^I > \mathbf{0}$,*

$$\begin{aligned} \text{Low received MUI: } & \frac{1}{w_i} \sum_{j \neq i} w_j \max_{1 \leq k \leq N} \left\{ \frac{|H_{ij}(k)|^2}{|H_{jj}(k)|^2} \cdot \text{innr}_{ij}(k) \right\} < 1, \quad \forall i = 1, \dots, I, \\ \text{Low generated MUI: } & \frac{1}{w_j} \sum_{i \neq j} w_i \max_{1 \leq k \leq N} \left\{ \frac{|H_{ij}(k)|^2}{|H_{jj}(k)|^2} \cdot \text{innr}_{ij}(k) \right\} < 1, \quad \forall j = 1, \dots, I \end{aligned} \quad (61)$$

Similar sufficient conditions can be obtained for $\mathbf{J}\mathbf{G}_{\text{low}}(k)$ to be positive semidefinite.

These conditions have an interesting physical interpretation: the uniqueness of the NE is ensured if the interference among the SUs is sufficiently small; this is clearly shown by Corollary 32. Specifically, the first condition in (61) can be interpreted as a constraint on the maximum amount of interference that each

receiver can tolerate, whereas the second condition introduces an upper bound on the maximum level of interference that each transmitter is allowed to generate. We will show shortly that these conditions play a role also in the convergence of the proposed distributed iterative algorithms. Moreover, depending on the level of interference in the network, the NEP $\mathcal{G}_{\text{siso}}$ is a P_{Υ} or monotone NEP, implying different properties and solution schemes of the game, as described next; we classify these two scenarios as *low-interference* and *medium/high-interference* regime, respectively.

The case of P_{Υ} NEP (low-interference regime). When the matrix $\Upsilon_{\mathbf{G}}$ is a P matrix (or positive definite), the NEP $\mathcal{G}_{\text{siso}}$ is a P_{Υ} NEP [Proposition 31 (c)]. Invoking Theorem 13, the unique NE of the game can be computed with convergence guarantee using Algorithm 1 on $\mathcal{G}_{\text{siso}}$, as stated in the next theorem.

Theorem 33 *Suppose that $\mathcal{G}_{\text{siso}}$ is a P_{Υ} real NEP. Then, any sequence $\{\mathbf{p}^{(n)}\}_{n=0}^{\infty}$ generated by Algorithm 1 applied to $\mathcal{G}_{\text{siso}}$ converges to the unique NE of the NEP, for any given updating schedule of the players satisfying assumptions A1)-A3).*

When implementing Algorithm 1, each user needs to compute his best-response solution, given the interference generated by the others. In Sec. 7.1.1 (cf. Lemma 35), we prove that the best-response for the game $\mathcal{G}_{\text{siso}}$ has a multi-level waterfilling-like expression, implying that each user can compute his optimal solution locally and very efficiently (he only needs to measure the overall MUI experienced at his receiver and “waterfill” over it). Therefore, Algorithm 1 results to be totally distributed and computationally efficient, which makes it appealing for practical implementation in CR scenarios.

The case of monotone NEP (medium/high-interference regime). When $\mathbf{J}\mathbf{G}_{\text{low}}$ is positive semidefinite, the NEP $\mathcal{G}_{\text{siso}}$ is a monotone NEP, having in general multiple solutions. In such a case, to compute a solution of $\mathcal{G}_{\text{siso}}$ with convergence guarantee, there are two available options, namely: PDAs (either in their exact or inexact form) and PTRAs. The former are the only feasible choice when the SUs are not willing to cooperate; whereas the latter requires some (albeit very limited) signaling among the SUs in favor of better performance (the algorithm converges to the “best” available solution). To the best of our knowledge, the above algorithms are in the signal processing and communication literature the first example of distributed power control schemes that converge *even in the presence of multiple Nash equilibria*. Note that, in all the aforementioned algorithms, the best-response of the SUs can be efficiently computed via a multi-level waterfilling expression; we address this issue in Sec. 7.1.1 below. We provide next an instance of the PTRAs along with its convergence conditions; PDAs are obtained as special cases of the PTRAs, we thus omit their description.

Equilibrium Selection via Proximal-Tikhonov Regularization Algorithm. According to the developments of Sec. 5.3, the first step is to choose a merit function that quantifies the quality of a NE of $\mathcal{G}_{\text{siso}}$. Different heuristics can be used; as an example, here we focus on the following merit function: given the vector $\mathbf{w} \triangleq (w_i)_{i=1}^I \geq 0$, let

$$\phi(\mathbf{p}) \triangleq \sum_{i=1}^I w_i \sum_{j \neq i} \sum_{k=1}^N |H_{ij}(k)|^2 p_i(k). \quad (62)$$

This choice is motivated by the intuition that among all the Nash equilibria of \mathcal{G} , a good candidate is the one that minimizes the overall interference among the users, measured by the function $\phi(\mathbf{p})$, likely resulting in an “higher” sum-rate $\sum_{i=1}^I r_i(\mathbf{p})$. The NE selection problem based on the merit function ϕ can be then formulated as:

$$\begin{aligned} & \underset{\mathbf{p}}{\text{minimize}} && \phi(\mathbf{p}) \\ & \text{subject to} && \mathbf{p} \in \text{SOL}(\mathcal{P}^{\text{siso}}, \mathbf{r}). \end{aligned} \quad (63)$$

Problem (63) is an instance of (33); we can then solve it by applying Algorithm 4 introduced in Sec. 5.3; which corresponds to solving a sequence of perturbed $P_{\mathbf{r}}$ NEPs denoted by $\mathcal{G}_{\tau, \varepsilon^{(n)}, \bar{\mathbf{p}}} = \langle \mathcal{P}^{\text{siso}}, (-r_i(\mathbf{p}) + \varepsilon^{(n)} \gamma_i^T \mathbf{p}_i + \frac{\tau}{2} \|\mathbf{p}_i - \bar{\mathbf{p}}_i\|^2)_{i=1}^I \rangle$, wherein each player i solves the following optimization problem: given \mathbf{p}_{-i} , $\bar{\mathbf{p}}$, and $\varepsilon^{(n)} > 0$,

$$\underset{\mathbf{p}_i \in \mathcal{P}_i^{\text{siso}}}{\text{maximize}} \quad r_i(\mathbf{p}_i, \mathbf{p}_{-i}) - \varepsilon^{(n)} \gamma_i^T \mathbf{p}_i - \frac{\tau}{2} \|\mathbf{p}_i - \bar{\mathbf{p}}_i\|^2 \quad (64)$$

where $\gamma \triangleq (\gamma_i)_{i=1}^I$, with each $\gamma_i \triangleq \left(\sum_{j \neq i} w_j |H_{ji}(k)|^2 \right)_{k=1}^N$. A partially asynchronous version of Algorithm 4 applied to (63) is described in Algorithm 5 below, and its convergence conditions are given in Theorem 34, which is a direct application of results in Theorem 13 and Theorem 22.

Algorithm 5: NE selection for the real NEP $\mathcal{G}_{\text{siso}}$

(Data) : $\{\varepsilon^{(n)}\} \downarrow 0$ and $\tau > 0$.

(S.0) : Choose any $\mathbf{p}^{(0)} \in \mathcal{P}^{\text{siso}}$ and a center $\bar{\mathbf{p}} \geq \mathbf{0}$ of the regularization; set $\bar{\varepsilon} = \varepsilon^{(0)}$ and $n = 0$.

(S.1) : If $\mathbf{p}^{(n)}$ satisfies a suitable termination criterion, STOP.

(S.2) : For each $i = 1, \dots, I$, compute $\mathbf{p}_i^{(n+1)}$ as

$$\mathbf{p}_i^{(n+1)} = \begin{cases} \mathbf{p}_i^* \in \underset{\mathbf{p}_i \in \mathcal{P}_i^{\text{siso}}}{\text{argmax}} \left\{ r_i \left(\mathbf{p}_i, \mathbf{p}_{-i}^{(\tau_{-i}(n))} \right) - \bar{\varepsilon} \gamma_i^T \mathbf{p}_i - \frac{\tau}{2} \|\mathbf{p}_i - \bar{\mathbf{p}}_i\|^2 \right\}, & \text{if } n \in \mathcal{T}_i \\ \mathbf{p}_i^{(n)}, & \text{otherwise} \end{cases} \quad (65)$$

(S.3) : If $\mathbf{p}^{(n+1)}$ is a NE of $\mathcal{G}_{\tau, \varepsilon^{(n)}, \bar{\mathbf{p}}}$, then update $\bar{\varepsilon}$ and the center $\bar{\mathbf{p}}$:

$$\bar{\varepsilon} = \varepsilon^{(n+1)} \quad \text{and} \quad \bar{\mathbf{p}}_i = \mathbf{p}_i^{(n+1)} \quad \forall i = 1, \dots, I; \quad (66)$$

(S.4) : $n \leftarrow n + 1$ and return to (S.1).

Theorem 34 *Given the optimization problem (63), suppose that: i) $\mathcal{G}_{\text{siso}}$ is a monotone NEP; ii) $\{\varepsilon^{(n)}\}$ is such that $\varepsilon^{(n)} \rightarrow 0$ and $\sum_{n=0}^{\infty} \varepsilon^{(n)} = \infty$; and iii) τ is chosen so that $\mathbf{\Upsilon}_{\mathbf{G}} + \tau \mathbf{I}$ is a P matrix. Then, the sequence $\{\mathbf{p}^{(n)}\}_{n=0}^{\infty}$ generated by Algorithm 5 has a limit point and every such limit point is a solution of (63).*

A sufficient condition for matrix $\mathbf{\Upsilon}_{\mathbf{G}} + \tau \mathbf{I}$ in Theorem 34 to be P is

$$\tau > \max_{1 \leq i \leq I} \left\{ \sum_{j \neq i} \max_{1 \leq k \leq N} \left\{ \frac{|H_{ij}(k)|^2}{|H_{ii}(k)|^2} \text{innr}_{ij}(k) \right\} \right\} - 1. \quad (67)$$

Algorithm 5 shows that, in the presence of multiple equilibria, one can still have converge even when best-response based schemes (cf. Algorithm 1) fail, provided that the SUs play a “sequence” of games rather than a one-shot game; moreover, to reach the NE that minimizes the overall MUI among the users, the players’ objective functions need to be modified in order to contain an additional term—the linear term $\varepsilon^{(n)} \gamma_i^T \mathbf{p}_i$ —whose task is to “measure” on the way the quality of the solution that the algorithm is going to reach. Such a term has also a physical interpretation: it represents a punishment imposed to the users for using too much power and thus generating too much MUI.

Note that the computation of the optimal power allocations of the users in Algorithm 5 can be performed locally and distributively by the users as previously discussed for the $P_{\mathbf{r}}$ NEP, once $\varepsilon^{(n)}$ and γ_i are given. The computation of γ_i requires an estimate from each user i of the cross-channel between its transmitter and the receivers of all SUs being in the coverage radius of user i . This estimate needs to be computed only once (before running the algorithm) and updated at the rate of the coherence time of the channel. When the computation of γ_i is not possible, one can still use Algorithm 5, setting $\gamma_i = \mathbf{0}$ in (65), which corresponds to solving the optimization problem (63) with $\phi(\mathbf{p}) = 0$, and thus computing just *one* of the solutions of the game $\mathcal{G}_{\text{siso}}$; Theorem 34 still guarantees convergence of the algorithm, even in the presence of multiple equilibria. The lack of any coordination however does not allow to perform any solution selection.

7.1.1 Efficient computation of the SISO proximal best-response solutions

In the distributed algorithms described in the previous section, each SU needs to compute the best-response solution of his rate maximization problem; see, e.g., (65). In this section, we provide an efficient method for computing such a solution. Motivated also by other resource allocation problems, such as [59], we introduce next a very general formulation that encompasses the optimization problems studied in this paper, whose optimal solution is proved to have a multi-level waterfilling-like expression, and provide an efficient algorithm to compute the optimal water-levels (dual variables).

Consider a parallel additive colored Gaussian-noise channel composed of N subchannels with channel gains $\{H_k\}_{k=1}^N$, subject to some spectral mask constraints $\mathbf{p}^{\max} = (p_k^{\max})_{k=1}^N$ and to several weighted average power constraints across the subchannels $\sum_{k=1}^N \mathbf{w}_k p_k \leq \boldsymbol{\alpha}$, where the inequality has to be intended component-wise and $\boldsymbol{\alpha} \geq \mathbf{0}$. Under this setup, the mathematical formulation of the optimization problem in the presence of given exogenous prices $\boldsymbol{\lambda} = (\lambda_k)_{k=1}^N$ and the proximal regularization in the objective is the following: given $\tau > 0$ and $\mathbf{c} = (c_k)_{k=1}^N$,

$$\begin{aligned} \underset{\mathbf{p}}{\text{maximize}} \quad & \sum_{k=1}^N [\log(1 + H_k p_k) - \lambda_k p_k] - \frac{\tau}{2} \|\mathbf{p} - \mathbf{c}\|^2 \\ & \sum_{k=1}^N \mathbf{w}_k p_k \leq \boldsymbol{\alpha} \\ & \mathbf{0} \leq \mathbf{p} \leq \mathbf{p}^{\max} \end{aligned} \tag{68}$$

where we make the following trivial assumptions without loss of generality: $\mathbf{0} < (H_k)_{k=1}^N < \infty$; $\mathbf{0} \leq \mathbf{w}_k \triangleq (w_{ki})_{i=1}^{N_c} < \infty$ for all $k = 1, \dots, N$, and linearly independent; $\mathbf{0} \leq (\lambda_k)_{k=1}^N < \infty$; $\mathbf{0} \leq (c_k)_{k=1}^N < \infty$; $\mathbf{0} < \mathbf{p}^{\max} < \infty$; and $\sum_{k=1}^N \mathbf{w}_k p_k^{\max} > \boldsymbol{\alpha}$.

Problem (68) is a convex problem with a polyhedral feasible set; the KKT are then necessary and sufficient conditions for the optimality. Solving the KKT system leads to the following structure for the optimal solution \mathbf{p}^* : denoting by $\boldsymbol{\mu} = (\mu_i)_{i=1}^{N_c} \geq \mathbf{0}$ the multipliers associated to the weighted average power constraints, for

each $k = 1, \dots, N$, we have

$$p_k^* = \begin{cases} p_k^{\max}, & \text{if } \lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k \leq \frac{H_k}{1+H_k p_k^{\max}} - \tau (p_k^{\max} - c_k) \\ p_k^* : \lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k + \tau (p_k^* - c_k) = \frac{H_k}{1+H_k p_k^*}, & \text{if } \frac{H_k}{1+H_k p_k^{\max}} - \tau (p_k^{\max} - c_k) < \lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k < H_k + \tau c_k \\ 0, & \text{if } H_k + \tau c_k \leq \lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k, \end{cases} \quad (69)$$

where the multipliers $\boldsymbol{\mu}$ have to be chosen to satisfy the complementarity conditions

$$0 \leq \mu_i \perp \alpha_i - \sum_{k=1}^N w_{ki} p_k^* \geq 0, \quad i = 1, \dots, N_c. \quad (70)$$

The optimal power allocation in (69) is defined only implicitly, as a solution of a quadratic equation. To obtain an explicit and simpler expression, we study first the properties of such equation.

Lemma 35 *Consider the quadratic equation $\tilde{\mu}_k + \tau (p_k - c_k) = \frac{H_k}{1+H_k p_k}$ in the variable p_k . The equation has the following properties: i) both roots are real, one root is always negative, and one root is always nonnegative; ii) both roots are decreasing in $\tilde{\mu}_k$, and iii) the nonnegative root is given by*

$$p_k^* = \frac{1}{2} \left(c_k - \frac{1}{H_k} \right) - \frac{1}{2\tau} \left[\tilde{\mu}_k - \sqrt{\left[\tilde{\mu}_k - \tau \left(c_k + \frac{1}{H_k} \right) \right]^2 + 4\tau} \right] \quad (71)$$

Proof. See Appendix H. ■

Applying now Lemma 35 to (69) [using, in particular, the fact that the positive root of the quadratic equation in (69) is decreasing in the term $\tilde{\mu}_k \triangleq \lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k$], it is not difficult to see that the optimal power allocation (69) can be equivalently written as

$$p_k^* = \left[\frac{1}{2} \left(c_k - \frac{1}{H_k} \right) - \frac{1}{2\tau} \left[\tilde{\mu}_k - \sqrt{\left[\tilde{\mu}_k - \tau \left(c_k + \frac{1}{H_k} \right) \right]^2 + 4\tau} \right] \right]_0^{p_k^{\max}} \quad k = 1, \dots, N \quad (72)$$

where each $\tilde{\mu}_k \triangleq \lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k$ and the water-level vector $\boldsymbol{\mu}$ has to be chosen to satisfy the complementarity conditions in (70), and $[x]_a^b$ denotes the Euclidean projection of x onto $[a, b]$, i.e., $[x]_a^b \triangleq \max(a, \min(x, b))$.

It is instructive to verify that the expression in (72) particularizes to the multi-level waterfilling when $\tau = 0$:

$$\lim_{\tau \rightarrow 0} \left[\frac{1}{2} \left(c_k - \frac{1}{H_k} \right) - \frac{1}{2\tau} \left[\tilde{\mu}_k - \sqrt{\left[\tilde{\mu}_k - \tau \left(c_k + \frac{1}{H_k} \right) \right]^2 + 4\tau} \right] \right]_0^{p_k^{\max}} = \left[\frac{1}{\tilde{\mu}_k} - \frac{1}{H_k} \right]_0^{p_k^{\max}}.$$

Also, when τ goes to infinity, the optimal power allocation is equal to the center of the regularization:

$$\lim_{\tau \rightarrow \infty} \left[\frac{1}{2} \left(c_k - \frac{1}{H_k} \right) - \frac{1}{2\tau} \left[\tilde{\mu}_k - \sqrt{\left[\tilde{\mu}_k - \tau \left(c_k + \frac{1}{H_k} \right) \right]^2 + 4\tau} \right] \right]_0^{p_k^{\max}} = [c_k]_0^{p_k^{\max}}.$$

The numerical computation of the water-level $\boldsymbol{\mu}$ in (72) so that the complementarity conditions in (70) are satisfied can be done efficiently in practice with a multiple nested bisection method as described in detail in Algorithm 6 below.

Algorithm 6: Multiple nested bisection algorithm for the computation of the proximal best-response in (72).

- (S.0) : Choose some accuracy ϵ .
- (S.1) : Set $\underline{\mu}_1 = 0$ and $\bar{\mu}_1 = \max_k \{(H_k - \tau c_k - \lambda_k) / w_{k1}\}$.
- (S.2) : Set $\mu_1 = (\underline{\mu}_1 + \bar{\mu}_1) / 2$.
- (S.3) : Solve for μ_2, \dots
- (S-2.1) : Set $\underline{\mu}_2 = 0$ and $\bar{\mu}_2 = \max_k \{(H_k - \tau c_k - \lambda_k - \sum_{j < 2} w_{kj} \mu_j) / w_{k2}\}$.
- (S-2.2) : Set $\mu_2 = (\underline{\mu}_2 + \bar{\mu}_2) / 2$.
- (S-2.3) : Solve for μ_3, \dots
- (S-3.1) : Set $\underline{\mu}_3 = 0$ and
- $\bar{\mu}_3 = \max_k \{(H_k - \tau c_k - \lambda_k - \sum_{j < 3} w_{kj} \mu_j) / w_{k3}\}$.
- (S-3.2) : Set $\mu_3 = (\underline{\mu}_3 + \bar{\mu}_3) / 2$.
- (S-3.3) : Solve for μ_4, \dots
- \vdots
- (S-3.4) : Using (72), if $\sum_{k=1}^N w_{k3} p_k^* < \alpha_3$ then set $\bar{\mu}_3 = \mu_3$, otherwise set $\underline{\mu}_3 = \mu_3$.
- (S-3.5) : If $\bar{\mu}_3 - \underline{\mu}_3 > \epsilon$, then go to (S-3.2).
- (S-2.4) : Using (72), if $\sum_{k=1}^N w_{k2} p_k^* < \alpha_2$ then set $\bar{\mu}_2 = \mu_2$, otherwise set $\underline{\mu}_2 = \mu_2$.
- (S-2.5) : If $\bar{\mu}_2 - \underline{\mu}_2 > \epsilon$, then go to (S-2.2).
- (S.4) : Using (72), if $\sum_{k=1}^N w_{k1} p_k^* < \alpha_1$ then set $\bar{\mu}_1 = \mu_1$, otherwise set $\underline{\mu}_1 = \mu_1$.
- (S.5) : If $\bar{\mu}_1 - \underline{\mu}_1 > \epsilon$ then go to (S.2); otherwise STOP.
-

The basic idea of the algorithm is to employ a bisection algorithm in μ_1 ; then, for a given μ_1 , use a bisection algorithm in μ_2 ; and then in μ_3 and so on. For the i th bisection level, the interval can be chosen as $[0, \max_k \{(H_k - \tau c_k - \lambda_k - \sum_{j < i} \mu_j w_{kj}) / w_{ki}\}]$. The convergence of the nested bisection method follows from the following nontrivial monotonicity result.

Lemma 36 *For a given value of $\boldsymbol{\mu}_{1:i} \triangleq [\mu_1, \dots, \mu_i]$, let $p_k^*(\boldsymbol{\mu}_{1:i})$ denote the optimal power allocation satisfying (72) and the complementarity conditions in (70) associated to the dual variables $\boldsymbol{\mu}_{i+1:N_c}(\boldsymbol{\mu}_{1:i})$. Then, the weighted sum power $\sum_{k=1}^N w_{ki} p_k^*(\boldsymbol{\mu}_{1:i})$ is monotonic decreasing in μ_i for fixed $\boldsymbol{\mu}_{1:i-1}$.*

Proof. See Appendix H. ■

We can now state our main result for the computation of the multi-level waterfilling solution in (72).

Proposition 37 *Algorithm 6 converges in no more than*

$$\prod_i \left\lceil \log_2 \left(\max_k \{(H_k - \tau c_k - \lambda_k) / w_{ki}\} / \epsilon \right) \right\rceil$$

iterations, where ϵ is the desired accuracy in the computation of the parameter $\boldsymbol{\mu}$.

Proof. Note that it is implicitly assumed that ϵ is sufficiently small such that the evaluation of the log is nonnegative. The convergence of the algorithm follows from the following fact: For fixed $\boldsymbol{\mu}_{1:i-1}$, the i th

bisection loop converges in at most $\left\lceil \log_2 \left(\max_k \left\{ \left(H_k - \tau c_k - \lambda_k - \sum_{j < i} \mu_j w_{kj} \right) / w_{ki} \right\} \right) / \epsilon \right\rceil$ iterations due to the monotonicity of $\sum_{k=1}^N w_{ki} p_k^*(\mu_{1:i})$ in μ_i , as given in Lemma 36. The worst case number of iterations for the i th bisection loop can be easily obtained by taking $\mu_j = 0$, for all $j < i$. ■

7.2 The MIMO case

Let us consider now the MIMO games. Because of space limitation, we focus only on the complex NEP $\mathcal{G}_{\text{mimo}}$ in (11); the analysis of the complex NEP $\mathcal{G}_{\text{femto}}$ in (13) follows the same lines, and thus is omitted.

Given $\mathcal{G}_{\text{mimo}}$, let us assume w.l.o.g. that all the matrices \mathbf{Q}_i in the game have the same dimensions $n_T \times n_T$. It follows from Proposition 24 (see also Example 25) that $\mathcal{G}_{\text{mimo}}$ is equivalent to the complex VI($\mathcal{P}^{\text{mimo}}, \mathbf{F}^{\mathbb{C}}$) where $\mathbf{F}^{\mathbb{C}}(\mathbf{Q}) \triangleq (\mathbf{F}_i^{\mathbb{C}}(\mathbf{Q}))_{i=1}^I : \mathcal{P}^{\text{mimo}} \rightarrow \mathbb{C}^{Q_{n_T \times n_T}}$, with each

$$\mathbf{F}_i^{\mathbb{C}}(\mathbf{Q}) \triangleq -(\nabla_{\mathbf{Q}_i} R_i(\mathbf{Q}))^* = -\mathbf{H}_{ii}^H \left(\mathbf{R}_{n_i} + \sum_{j=1}^I \mathbf{H}_{ij} \mathbf{Q}_j \mathbf{H}_{ij}^H \right)^{-1} \mathbf{H}_{ii}. \quad (73)$$

According to Proposition 28, the monotonicity/P properties of $\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$ on $\mathcal{P}^{\text{mimo}}$ are related to the properties of the augmented Jacobian matrix $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$. To obtain sufficient conditions easy to be checked, let us introduce the following $I \times I$ matrix $\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}} \in \mathbb{R}^{I \times I}$ obtained by properly “condensing” $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$ (see Appendix I), and defined as (we implicitly assume that all the channel matrices \mathbf{H}_{ii} are full column rank):

$$[\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}]_{ij} \triangleq \begin{cases} 1 & \text{if } i = j \\ -\rho \left(\mathbf{H}_{ii}^{\dagger H} \mathbf{H}_{ij}^H \mathbf{H}_{ij} \mathbf{H}_{ii}^{\dagger} \right) \cdot \text{INNR}_{ij} & \text{if } i \neq j, \end{cases} \quad (74)$$

where \mathbf{A}^{\dagger} denotes the Moore–Penrose pseudoinverse of \mathbf{A} , $\rho(\mathbf{A})$ is the spectral radius of \mathbf{A} , and INNR_{ij} is defined as

$$\text{INNR}_{ij} \triangleq \frac{\rho \left(\mathbf{R}_{n_i} + \sum_{j=1}^I P_j \mathbf{H}_{ij} \mathbf{H}_{ij}^H \right)}{\lambda_{\text{least}}(\mathbf{R}_{n_i})} \quad (75)$$

where, in this case, $\lambda_{\text{least}}(\mathbf{R}_{n_i})$ is the minimum eigenvalue of the (positive definite) matrix \mathbf{R}_{n_i} . Note that since \mathbf{H}_{ii} are full column rank, we have $\mathbf{H}_{ii}^{\dagger} = (\mathbf{H}_{ii}^H \mathbf{H}_{ii})^{-1} \mathbf{H}_{ii}^H$.

Using the above definitions, Proposition 24 and Proposition 28, we obtain the following characterization for the complex game $\mathcal{G}_{\text{mimo}}$.

Proposition 38 *Given the NEP $\mathcal{G}_{\text{mimo}} = \langle \mathcal{P}^{\text{mimo}}, (R_i)_{i=1}^I \rangle$, the following hold.*

- (a) $\mathcal{G}_{\text{mimo}}$ is equivalent to the complex VI($\mathcal{P}^{\text{mimo}}, \mathbf{F}^{\mathbb{C}}$), which has a nonempty and compact solution set;
- (b) Suppose that matrix $\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ is positive semidefinite. Then $\mathbf{F}^{\mathbb{C}}$ is monotone on $\mathcal{P}^{\text{mimo}}$; therefore, $\mathcal{G}_{\text{mimo}}$ is a monotone complex NEP;
- (c) Suppose that $\mathbf{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ is a P -matrix (positive definite matrix). Then $\mathbf{F}^{\mathbb{C}}$ is a uniformly P -function (strongly monotone function) on $\mathcal{P}^{\text{mimo}}$; therefore, $\mathcal{G}_{\text{mimo}}$ is a complex $P_{\mathbf{\Upsilon}}$ NEP and has a unique NE.

Proof. See Appendix I. ■

Corollary 39 *The matrix $\Upsilon_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ is a P (or positive definite) matrix if one (or both) of the following two sets of conditions are satisfied: for some $\mathbf{w} = (w_i)_{i=1}^I > \mathbf{0}$,*

$$\begin{aligned} \text{Low received MUI:} \quad & \frac{1}{w_i} \sum_{j \neq i} w_j \left\{ \rho \left(\mathbf{H}_{ii}^{\dagger H} \mathbf{H}_{ij}^H \mathbf{H}_{ij} \mathbf{H}_{ii}^{\dagger} \right) \cdot \text{INNR}_{ij} \right\} < 1, \quad \forall i = 1, \dots, I \\ \text{Low generated MUI:} \quad & \frac{1}{w_j} \sum_{i \neq j} w_i \left\{ \rho \left(\mathbf{H}_{ii}^{\dagger H} \mathbf{H}_{ij}^H \mathbf{H}_{ij} \mathbf{H}_{ii}^{\dagger} \right) \cdot \text{INNR}_{ij} \right\} < 1, \quad \forall j = 1, \dots, I \end{aligned} \quad (76)$$

It is interesting to observe that conditions for $\mathbf{F}^{\mathbb{C}}$ to be a uniformly P-function on $\mathcal{P}^{\text{mimo}}$ are the natural generalization of those obtained for $\mathcal{G}_{\text{siso}}$ to be a P_{Υ} game; they thus have the same physical interpretation, for which we refer the reader to Sec. 7.1. Based on that, in Proposition 38 we used the same terminology as in Definition 11, namely: $\mathcal{G}_{\text{mimo}}$ is a complex P_{Υ} NEP if $\Upsilon_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ is a P matrix, whereas is a complex monotone NEP if $\Upsilon_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ is a semidefinite matrix. For these two classes of NEPs we can devise distributed algorithms having the same convergence properties and features of those developed in Sec. 5.1 and Sec. 5.2 for real P_{Υ} and monotone NEPs, respectively. The main results are briefly listed next; proofs are based on the same techniques used to prove Lemma 24 and Proposition 38, and thus are omitted.

The case of P_{Υ} NEP $\mathcal{G}_{\text{mimo}}$ (low-interference regime). When the matrix $\Upsilon_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ is a P matrix, the unique NE of the game can be computed using Algorithm 1 on $\mathcal{G}_{\text{mimo}}$, as stated in the next theorem.

Theorem 40 *Suppose that $\mathcal{G}_{\text{mimo}}$ is a complex P_{Υ} NEP. Then, any sequence $\{\mathbf{Q}^{(n)}\}_{n=0}^{\infty}$ generated by Algorithm 1 applied to $\mathcal{G}_{\text{mimo}}$ converges to the unique NE of the NEP, for any given updating schedule of the players satisfying assumptions A1)-A3).*

The algorithm has the same desired features as the one obtained for the SISO case; see Sec. 7.1.

The case of monotone NEP $\mathcal{G}_{\text{mimo}}$ (medium-interference regime). When $\Upsilon_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}} \succeq \mathbf{0}$, $\mathcal{G}_{\text{mimo}}$ is a monotone complex NEP; in the presence of multiple solutions, we need to choose first a merit function quantifying the quality of a NE of $\mathcal{G}_{\text{mimo}}$; similarly to the SISO case, we consider the overall interference generated by all the SUs:

$$\phi(\mathbf{Q}) \triangleq \sum_{i=1}^I w_i \sum_{j \neq i} \text{tr}(\mathbf{H}_{ij} \mathbf{Q}_j \mathbf{H}_{ij}^H), \quad (77)$$

where w_i 's are positive given weights. Building on the equivalence between the game $\mathcal{G}_{\text{mimo}}$ and the VI($\mathcal{P}^{\text{mimo}}, \mathbf{F}^{\mathbb{C}}$) (Lemma 24), the NE selection problem based on the merit function ϕ becomes:

$$\begin{aligned} & \underset{\mathbf{Q}}{\text{minimize}} \quad \phi(\mathbf{Q}) \\ & \text{subject to} \quad \mathbf{Q} \in \text{SOL}_{\mathbb{C}}(\mathcal{P}^{\text{mimo}}, \mathbf{F}^{\mathbb{C}}). \end{aligned} \quad (78)$$

A solution of (78) can be computed in a distributed way using Algorithm 4 applied to (78); it can be shown that this corresponds to solving a sequence of perturbed complex NEPs denoted by $\mathcal{G}_{\tau, \varepsilon^{(n)}, \bar{\mathbf{Q}}} \triangleq \left\langle \mathcal{P}^{\text{mimo}}, \left(-R_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) + \varepsilon^{(n)} \langle \bullet, \mathbf{\Pi}_i \rangle + \frac{\tau}{2} \|\bullet - \bar{\mathbf{Q}}_i\|_F^2 \right)_{i=1}^I \right\rangle$, where each player solves the following convex optimization problem: given $\mathbf{Q}_{-i} \succeq \mathbf{0}$, $\bar{\mathbf{Q}}_i \in \mathbb{C}^{n \times n}$, $\tau > 0$, $\varepsilon^{(n)} > 0$, and $\mathbf{\Pi}_i \triangleq \sum_{j \neq i} w_j \mathbf{H}_{ji}^H \mathbf{H}_{ji}$,

$$\underset{\mathbf{Q}_i \in \mathcal{P}_i^{\text{mimo}}}{\text{maximize}} \quad R_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) - \varepsilon^{(n)} \langle \mathbf{Q}_i, \mathbf{\Pi}_i \rangle - \frac{\tau}{2} \|\mathbf{Q}_i - \bar{\mathbf{Q}}_i\|_F^2. \quad (79)$$

The formal description of the algorithm along with its convergence properties are given in Algorithm 7 and Theorem 41 below, respectively.

Algorithm 7: NE selection for the complex NEP $\mathcal{G}_{\text{mimo}}$

(Data) : $\{\varepsilon^{(n)}\} \downarrow 0$ and $\tau > 0$.

(S.0) : Choose any $\mathbf{Q}^{(0)} \in \mathcal{P}^{\text{mimo}}$ and a center of the regularization $\bar{\mathbf{Q}} \succeq \mathbf{0}$, set $\bar{\varepsilon} = \varepsilon^{(0)}$ and $n = 0$.

(S.1) : If $\mathbf{Q}^{(n)}$ satisfies a suitable termination criterion, STOP.

(S.2) : For each $i = 1, \dots, I$, compute $\mathbf{Q}_i^{(n+1)}$ as

$$\mathbf{Q}_i^{(n+1)} = \begin{cases} \mathbf{Q}_i^* \in \operatorname{argmax}_{\mathbf{Q}_i \in \mathcal{P}_i^{\text{mimo}}} \left\{ R_i \left(\mathbf{Q}_i, \mathbf{Q}_{-i}^{(\tau - i(n))} \right) - \bar{\varepsilon} \langle \mathbf{Q}_i, \mathbf{\Pi}_i \rangle - \frac{\tau}{2} \|\mathbf{Q}_i - \bar{\mathbf{Q}}_i\|_F^2 \right\}, & \text{if } n \in \mathcal{T}_i \\ \mathbf{Q}_i^{(n)}, & \text{otherwise} \end{cases} \quad (80)$$

(S.3) : If $\mathbf{Q}^{(n+1)}$ is a NE of $\mathcal{G}_{\alpha, \varepsilon^{(n)}, \bar{\mathbf{Q}}}$, then update $\bar{\varepsilon}$ and the center $\bar{\mathbf{Q}}$:

$$\bar{\varepsilon} = \varepsilon^{(n+1)} \quad \text{and} \quad \bar{\mathbf{Q}}_i = \mathbf{Q}_i^{(n+1)} \quad \forall i = 1, \dots, I; \quad (81)$$

(S.4) : $n \leftarrow n + 1$ and return to (S.1).

Theorem 41 *Given the optimization problem (78), suppose that: i) $\mathcal{G}_{\text{mimo}}$ is a complex monotone NEP; ii) $\{\varepsilon^{(n)}\}$ is such that $\varepsilon^{(n)} \rightarrow 0$ and $\sum_{n=0}^{\infty} \varepsilon^{(n)} = \infty$; and iii) τ is chosen so that $\mathbf{\Upsilon}_{\text{FC}}^{\text{mimo}} + \tau \mathbf{I}$ is a P matrix. Then, the sequence $\{\mathbf{Q}^{(n)}\}_{n=0}^{\infty}$ generated by Algorithm 7 has a limit point and every such point is a solution of (78).*

A sufficient condition for matrix $\mathbf{\Upsilon}_{\text{FC}}^{\text{mimo}} + \tau \mathbf{I}$ in Theorem 41 to be P is

$$\tau > \max_{1 \leq i \leq I} \left\{ \sum_{j \neq i} \rho \left(\mathbf{H}_{ii}^{\dagger H} \mathbf{H}_{ij}^H \mathbf{H}_{ij} \mathbf{H}_{ii}^{\dagger} \right) \cdot \text{INNR}_{ij} \right\} - 1. \quad (82)$$

Algorithm 7 has the same desirable properties of Algorithm 5 introduced in Sec. 7.1, which we refer to for a detailed discussion on its practical implementation. Here, we only observe that, in the scenarios where the SUs cannot compute the optimal pricing matrix $\mathbf{\Pi}_i$ we can still use Algorithm 7 setting $\mathbf{\Pi}_i = \mathbf{0}$ in (80), and converge (under assumptions of Theorem 41) to *any one* of the solutions of the game $\mathcal{G}_{\text{mimo}}$.

8 Numerical Results

In this section, we compare some of the proposed algorithms solving NEP $\mathcal{G}_{\text{siso}}$ in (7) and $\mathcal{G}_{\text{mimo}}$ in (11) in terms of achievable rates and convergence speed. We also compare the performance of our distributed schemes with those achievable computing a stationary solution of the related sum-rate optimization problem (under the same power and interference constraints). Among all the schemes available in the literature, we consider here the pricing-based algorithm proposed in [59] for the SISO case, and the solution methods proposed in [59, 60] for the MIMO setting; we slight modified the schemes in [59, 60] in order to include the interference

constraints we have proposed in this paper. Algorithms in [59, 60] are the benchmark methods for this kind of problems. It is important to remark that algorithms in [59, 60] are either centralized or require heavy signaling and coordination among the users to be implemented. The proposed comparison then sheds light on the trade-off between network signaling and system performance (measured in terms of the achievable sum-rate). Finally, we contrast our best-response schemes with gradient-response ones [34].

Example #1 (NE selection vs. no selection). In Figure 4, we plot the SUs' sum-rate $\sum_{i=1}^I r_i(\mathbf{p})$ versus the iteration index, achieved by the following algorithms applied to the NEP $\mathcal{G}_{\text{siso}}$: i) The Jacobi version of the proximal decomposition algorithm described in Algorithm 2 (green line curve); ii) the Jacobi version of Algorithm 5 (blue line curve), where $\phi(\mathbf{p})$ is given by (62); iii) the same Algorithm 5 applied to (63), where $\phi(\mathbf{p})$ in (62) is replaced by $-\phi(\mathbf{p})$ (red line curve); and iv) the Jacobi Dynamic Pricing-based Algorithm (DPA) reaching a stationary solution of the sum-rate maximization problem (black line curve) [59]. The choice of the merit function $-\phi(\mathbf{p})$ leads to the selection of the NE solution that maximizes the overall MUI in the system, which provides a benchmark of the sum-rate variability and an estimate of the worst-case performance over the set of the Nash equilibria of the game. Any best-response solution involved in the optimization problems is computed using the waterfilling-like expression introduced in Sec. 7.1.1.

We examined the behavior of the above algorithms under the following setup. We considered a CR network modeled as a Gaussian parallel interference channel, composed of $I = 25$ active users and two PUs; all the users are randomly placed within an hexagonal cell; the channels of all the links are simulated as FIR filter of order $L = 10$, where each tap is a zero mean complex Gaussian random variable with variance equal to $1/L$; the channel transfer functions are the FFT of the corresponding impulse responses over $N = 128$ points (carriers). We focused on two scenarios, namely low interference regime (corresponding to $\mathbf{\Upsilon}_{\mathbf{G}}$ being a P matrix) and high interference regime (corresponding to $\mathbf{J}\mathbf{G}_{\text{low}}$ being positive semidefinite). The thermal noise variance $\sigma_i^2(k)$ is set to one for all k and i , and the transmit power budget P_i is chosen so that the Signal-to-Noise-Ratio (SNR) of each user $\text{SNR}_i \triangleq 10 \log_{10} (P_i/\sigma_i^2(k)) = 5\text{dB}$ for all i and k . The PUs impose interference-temperature limit constraints; for the sake of simplicity, we assume the same interference thresholds $\alpha_i = \alpha \mathbf{1}$ for all the SUs, with $\alpha = 10^{-3}$ (this choice of α is such that the power budget constraints of the SUs are not active at any optimal solution). All the algorithms are initialized by the same starting point, chosen randomly in the set $\mathcal{P}^{\text{siso}}$, and are terminated when the Euclidean norm of the error in two consecutive iterations becomes smaller than 10^{-6} . In the PTDA, we chose the center $\bar{\mathbf{p}}$ of the regularization randomly in \mathcal{P} , $\tau = 3.5$, and $\varepsilon^{(n)} = \varepsilon^{(0)}/(1 + 10n)$, where $\varepsilon^{(0)} = 0.5$; in all the algorithms, the termination criterion of the inner loop, if any, is the same as the outer loop. The above choice of the free parameters is the result of some preliminary tests; it is important to remark however that the proposed algorithm has been observed to be robust against the variation of the aforementioned parameters.

The following comments are in order from Figure 4. In the case of multiple Nash equilibria, the sum-rate performance of the network can vary significantly over the set of the Nash equilibria; the relative sum-rate gap between the “worst” and “best” NE is more than 90%. As expected, Algorithm 5 outperforms Algorithm 2, which validates the use of criterion (62) in choosing the NE. Moreover the sum-rate loss with respect to the DPA is very limited, and more than acceptable if one takes into account that, to be implemented, the DPA requires a significant signaling among the users at each iteration. There are scenarios where such a signaling exchange is not feasible, because the users are heterogeneous systems that are not willing to cooperate with

each other; in all these cases the proposed algorithm is a good candidate: it outperforms Algorithm 2 while keeping the same distributed nature and guaranteeing convergence. When the NE of the game is unique [$\Upsilon_{\mathbf{G}}$ is a P matrix], as expected, both Algorithm 2 and Algorithm 5 converge to the same sum-rate solution. Interestingly, this solution seems to coincide also with the one achieved by the DPA. Finally, note that our algorithms converge quite fast.

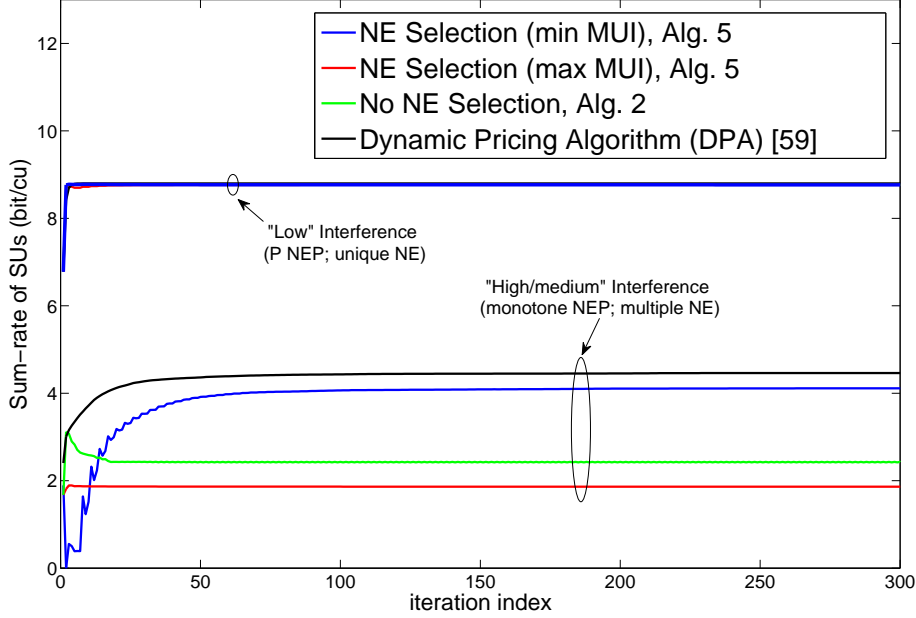


Figure 4: Comparison of distributed algorithms solving $\mathcal{G}_{\text{siso}}$: Sum-rate of the SUs versus the iteration index.

Figure 5 shows the average performance of the algorithms i)-iii) considered in Figure 4. We plotted the average sum-rate versus the SNR $\triangleq P$, with $P_i = P$ and $\sigma_i^2(k) = 1$ for all i and k , achievable at the NE reached by the Algorithm 2 and Algorithm 5. The curves are averages over 5000 random channel realization chosen so that the \mathbf{JG}_{low} is positive semidefinite. The rest of the parameters are the same as in Figure 4. Figure 5 confirms the superior performance of Algorithm 5 with respect to Algorithm 2 that does not perform any equilibrium selection.

Finally, it is worth observing that projection-response algorithms proposed in [34] and [30, 33] cannot be used to solve the $\mathbf{P}\Upsilon$ NEP $\mathcal{G}_{\text{siso}}$ (unless it is also monotone), even if the game has a unique NE.

Example #2 (Comparison with gradient-response algorithms for monotone VIs). In Figures 6 and 7 we compare some algorithms solving the game $\mathcal{G}_{\text{siso}}$ under the monotonicity assumption ($\mathcal{G}_{\text{siso}}$ is a monotone NEP); the setup is the same of Figure 4. More specifically, in Figure 6, we plot the SUs rates versus the iteration index achieved by Algorithm 2 and the Iterative Tikhonov Algorithm recently proposed in [34] for solving monotone VIs. In the latter algorithm we chose the variable step-size sequences $\gamma_n = n^{-0.4}$ and $\delta_n = n^{-0.49}$ so that (sufficient) conditions for the convergence given in [33, Prop. 15.1] are satisfied (we use the same notation as in [34]; see therein for the details). The figure clearly shows that our best-response-based scheme converges in a very few iterations, whereas the gradient-response algorithm needs much more iterations (two orders of magnitude more) to reach comparable performance. The same convergence properties

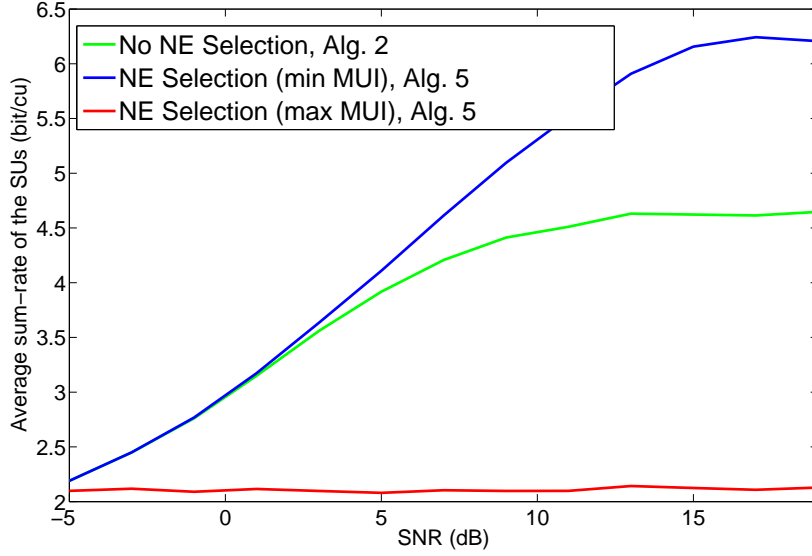


Figure 5: Average sum-rate of the SUs versus the SNR at the NE of \mathcal{G}_{iso} .

as in Figure 6 has been experienced for all the channel realizations we simulated.

Figure 7 shows an example where Algorithm 1 does not converge because of the high interference among the SUs (sufficient conditions in Theorem 33 are not satisfied), whereas Algorithm 2 still converges (in the sense of Theorem 19) in a few iterations. Interestingly, we experienced convergence of Algorithm 2 even when \mathbf{JG}_{low} is not positive semidefinite, provided that the error sequence $\{\varepsilon^{(n)}\}$ is properly chosen (e.g., $\{\varepsilon^{(n)}\}$ goes to zero sufficiently slow), showing that the conditions for the convergence are far from being necessary.

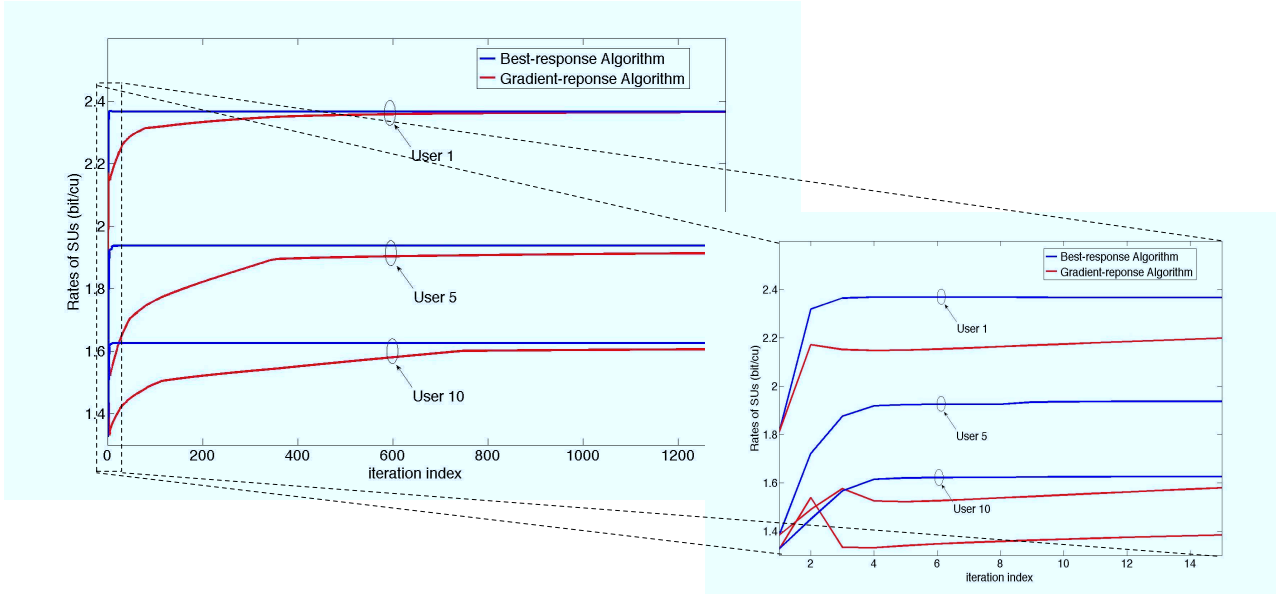


Figure 6: Typical behavior of gradient-response versus best-response algorithms solving the monotone NEP \mathcal{G}_{iso} : rate of three out of 25 users versus the iterations achievable by the gradient-response algorithm [34] (red-line curves) and the simultaneous best-response algorithm described in Algorithm 1 (blue-line curves).

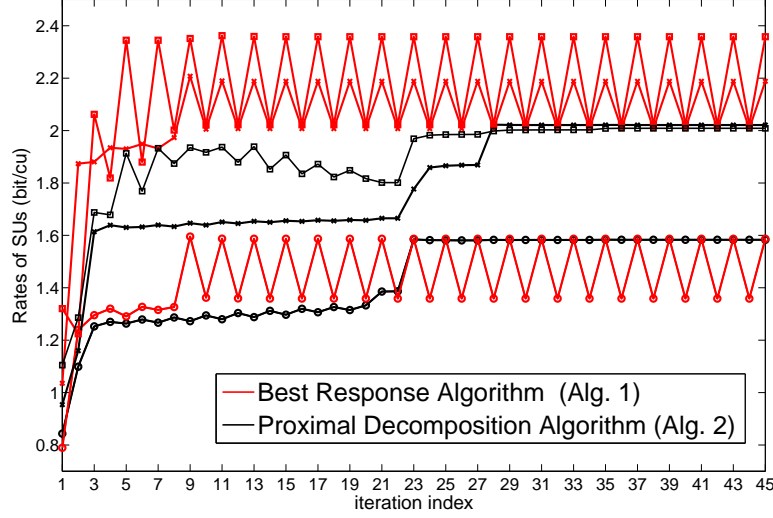


Figure 7: Comparison of distributed algorithms for the monotone NEP \mathcal{G}^{iso} : Rates of the SUs versus iterations achievable by GIWFA (Algorithm 1) and PDA (Algorithm 2).

Example #3 (NE selection vs. stationary solutions: the MIMO case). We compare here some of the proposed algorithms in the MIMO setting. We consider the same scenario as in Figure 1, with the only difference that now the transceivers $i = 1, \dots, I$ are equipped with $n_{T_i} = n_{R_i} = 3$ antennas and there are $I = 5$ active SUs. The channels are MIMO frequency-selective (the order of the channels is $L = 10$ and the number of subcarriers is $N = 128$) and are generated in order to guarantee that the matrix $\mathbf{\Upsilon}_{\mathbf{F}^c}$ in (74) is positive semidefinite, resulting thus in a monotone NEP $\mathcal{G}^{\text{mimo}}$. Soft average power shaping interference constraints are imposed to each SU along the direction of the primary transmitters (the column of the matrices \mathbf{G}_{pi} are the steering vectors representing the directions of the primary receivers $p = 1, 2$ as observed from the secondary transmitters $i = 1, \dots, I$); all the interference threshold are assumed to be equal and set to $I_{pi}^{\text{ave}} = 10^{-3}$. The best-response of each user cannot be computed in closed form (unless the proximal regularization is not included in the objective function), but can be efficiently computed using any nonlinear programming solvers (each player's optimization problem is convex). In Figure 8 we plot the sum-rate versus the iteration index achieved by the Jacobi version of Algorithm 2 (green-line curve), Algorithm 7 based on the merit function $\phi(\mathbf{Q})$ defined in (77) (blue line curve); and iii) the Gauss-Seidel based algorithm proposed in [60] to compute stationary solutions of the sum-rate problem (we adapted the algorithm in [60] including the interference constraints in the feasible set of the optimization problem). Figure 8 shows that trade-off between performance and signaling that is achievable by the three algorithms: Algorithm 7 implementing a NE selection leads to better sum-rates than Algorithm 2 at the cost of almost the same signaling among the SUs of classical MIMO IWFA's (a constant price depending on the cross-channel matrices needs to be computing by each SU before running the algorithms); higher sum-rates can be achieved using algorithm in [60] but at the cost of more signaling among the SUs, which is not admissible in all CR scenarios (note that in the MIMO case, the scheme [60] requires the SUs to exchange matrix informations at each iteration of the algorithm).

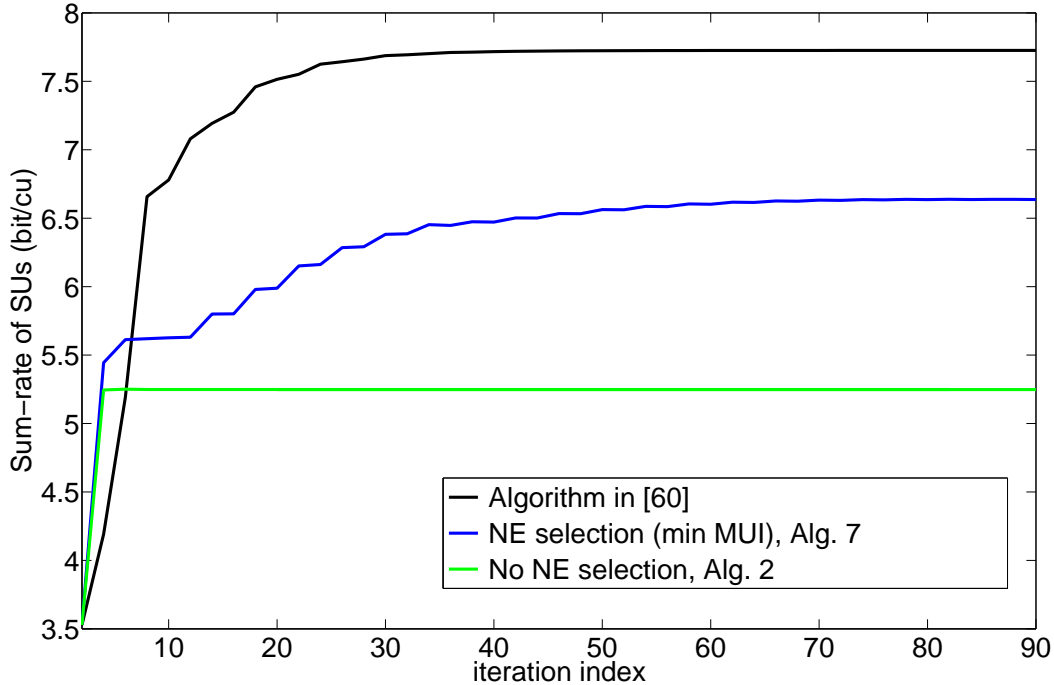


Figure 8: Comparison of distributed algorithms solving the game $\mathcal{G}_{\text{mimo}}$: Sum-rate of the SUs versus iterations.

9 Conclusions

In this paper, we have proposed a novel method based on VIs suitable to study and solve general real or complex player-convex NEPs, having (possibly) multiple solutions. The proposed framework has many desirable new features, such as: i) it can be applied to real and complex NEPs having no specific structure and for which the best-response mapping is not available in closed form or unique; ii) the algorithms proposed for computing a NE converge under mild conditions that do not imply the uniqueness of the equilibrium; and iii) in the presence of multiple equilibria, one can control the quality of the computed solution by guaranteeing convergence to the “best” NE (according to some prescribed criterion), at the cost of limited signaling among the players (or no signaling at all). These features make the proposed algorithms applicable to a variety of scenarios in different fields; the choice of one scheme with respect to the other will depend on the trade-off between signaling and performance that the users are willing to exchange/achieve. The analysis of algorithms for complex NEPs hinges on the definition of the VI problem in the complex domain; this new class of VIs along with their properties are introduced and studied for the first time in this paper.

Finally, to have suitable case studies, we applied the proposed framework to solve some novel NEPs modeling various resource allocation problems in SISO/MIMO CR and femtocell systems. The resulting distributed best-response algorithms were shown to converge even when current schemes proposed in the literature for related problems fail. Numerical results showed the superiority of our approach with respect to plain noncooperative solutions as well as good performance with respect to centralized solutions, in favor of very limited signaling among the players.

Appendix

A State-of-the-art Solution Analysis of Partitioned VI

We summarize herein the state-of-the-art results dealing with the solution analysis of partitioned VIs, which are more general than those given in Theorem 3; by Proposition 7, these results apply readily to player-convex NEPs. Note that some of them are drawn from [30], some others have not appeared in the literature and are new in their genre.

Before stating the main theorem, we introduce the following intermediate definitions. Given the partitioned VI(\mathcal{Q}, \mathbf{F}), let $\mathbf{F}_Q^{\text{nor}}$ denote the *normal map* of this VI, defined as

$$\mathbf{F}_Q^{\text{nor}}(\mathbf{z}) \triangleq \mathbf{F}(\Pi_Q(\mathbf{z})) + \mathbf{z} - \Pi_Q(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^n,$$

where $\Pi_Q(\mathbf{z})$ denotes the Euclidean projection of \mathbf{z} onto \mathcal{Q} . A solution \mathbf{x}^* of the VI(\mathcal{Q}, \mathbf{F}) is *isolated* if there exists an open neighborhood \mathcal{N} of \mathbf{x}^* such that $\mathcal{N} \cap \text{SOL}(\mathcal{Q}, \mathbf{F}) = \{\mathbf{x}^*\}$. Lastly, we let $\deg(\mathbf{F}_Q^{\text{nor}}, \mathcal{O})$ denote the *degree* of the map $\mathbf{F}_Q^{\text{nor}}$ with respect to the bounded open set \mathcal{O} , provided that this degree is well defined (see [30, Def. 2.1.1] for a formal definition of the degree of a map).

Theorem 42 Let $\mathcal{Q} \triangleq \prod_{i=1}^I \mathcal{Q}_i$ with each $\mathcal{Q}_i \subseteq \mathbb{R}^{n_i}$ being closed and convex (albeit not necessarily bounded).

Let $\mathbf{F} = (\mathbf{F}_i(\mathbf{x}))_{i=1}^I : \mathcal{Q} \rightarrow \mathbb{R}^n$ be a continuous P_0 function on \mathcal{Q} . Then, the following statements hold:

- (a) **Existence of a solution** ([30, Th. 3.5.11]): $\text{SOL}(\mathcal{Q}, \mathbf{F})$ is nonempty if and only if there exists a $\mathbf{x}^{\text{ref}} = (\mathbf{x}_i^{\text{ref}})_{i=1}^I \in \mathcal{Q}$ such that the set

$$L_{<} \triangleq \left\{ \mathbf{x} = (\mathbf{x}_i)_{i=1}^I \in \mathcal{Q} \mid \mathbf{F}_i(\mathbf{x})^T (\mathbf{x}_i - \mathbf{x}_i^{\text{ref}}) < 0 \text{ for all } i \text{ such that } \mathbf{x}_i \neq \mathbf{x}_i^{\text{ref}} \right\} \quad (83)$$

is bounded. The “if” part holds true even if \mathbf{F} is not a P_0 function [30, Prop. 3.5.1];

- (b) **Boundedness of solutions** ([30, Th. 5.5.15]): $\text{SOL}(\mathcal{Q}, \mathbf{F})$ is nonempty and bounded if and only if there exists a bounded open set \mathcal{O} containing $(\mathbf{F}_Q^{\text{nor}})^{-1}(0)$ such that $\deg(\mathbf{F}_Q^{\text{nor}}, \mathcal{O})$ is nonzero. A sufficient condition for this to hold is the existence of a $\mathbf{x}^{\text{ref}} = (\mathbf{x}_i^{\text{ref}})_{i=1}^I \in \mathcal{Q}$ such that the set

$$L'_{\leq} \triangleq \left\{ \mathbf{x} = (\mathbf{x}_i)_{i=1}^I \in \mathcal{Q} \mid \max_{1 \leq i \leq I} \mathbf{F}_i(\mathbf{x})^T (\mathbf{x}_i - \mathbf{x}_i^{\text{ref}}) \leq 0 \right\} \quad (84)$$

is bounded. The boundedness of L'_{\leq} is a sufficient condition for $\text{SOL}(\mathcal{Q}, \mathbf{F})$ being nonempty and compact even if \mathbf{F} is not a P_0 function [30, Prop. 3.5.1].

- (c) **Connectedness of solutions** ([30, Th. 3.6.6(a)]): If $\text{SOL}(\mathcal{Q}, \mathbf{F})$ is nonempty and bounded, then it is connected;
- (d) **Uniqueness of the solution** ([30, Th. 3.6.6(b)]): The VI(\mathcal{Q}, \mathbf{F}) has a unique solution if and only if it has an isolated solution. \square

Theorem 42 is very general and provides the state-of-the-art conditions for the above properties of a solution of a partitioned VI; for instance, Theorem 3 comes as special case.

We conclude this section providing conditions for the isolatedness of a solution of a the partitioned VI (and thus the globally uniqueness); we also present necessary and sufficient conditions for the uniqueness of the solution of a monotone nonlinear complementarity problem. The latter conditions are new in their genre.

Sufficient conditions for the isolatedness of a solution to the partitioned VI can be obtained by introducing the critical cone at a solution $\mathbf{x}^* \in \text{SOL}(\mathcal{Q}, \mathbf{F})$; by definition, this cone is given by:

$$\mathcal{C}(\mathbf{x}^*; \mathcal{Q}, \mathbf{F}) \triangleq \prod_{i=1}^I \mathcal{C}(\mathbf{x}_i^*, \mathcal{Q}_i, \mathbf{F}_i) \quad \text{with} \quad \mathcal{C}(\mathbf{x}_i^*, \mathcal{Q}_i, \mathbf{F}_i) \triangleq \mathcal{T}(\mathcal{Q}_i; \mathbf{x}_i^*) \cap \mathbf{F}_i(\mathbf{x}^*)^\perp, \quad (85)$$

where $\mathcal{T}(\mathcal{Q}_i; \mathbf{x}_i^*)$ denotes the *tangent cone* of \mathcal{Q}_i at \mathbf{x}_i^* . Exploiting only the Cartesian structure of \mathcal{Q} , the result below does not require the P_0 -property of \mathbf{F} .

Proposition 43 *Let \mathbf{F} be continuously differentiable on \mathcal{K} . The following two statements hold for a solution $\mathbf{x}^* \in \text{SOL}(\mathcal{Q}, \mathbf{F})$:*

(a) *If*

$$\max_{1 \leq i \leq I} (\mathbf{u}_i)^T [\mathbf{J}_{\mathbf{x}} \mathbf{F}_i(\mathbf{x}^*) \mathbf{u}] > 0 \quad \text{for all nonzero } \mathbf{u} = (\mathbf{u}_i)_{i=1}^I \in \mathcal{C}(\mathbf{x}^*; \mathcal{Q}, \mathbf{F}),$$

then \mathbf{x}^ is an isolated zero of the VI(\mathcal{Q}, \mathbf{F}).*

(b) *If each \mathcal{Q}_i is polyhedral and if the complementarity problem:*

$$\mathcal{C}(\mathbf{x}^*; \mathcal{Q}, \mathbf{F}) \ni \mathbf{u} \perp \mathbf{J}_{\mathbf{x}} \mathbf{F}_i(\mathbf{x}^*) \mathbf{u} \in \mathcal{C}(\mathbf{x}^*; \mathcal{Q}, \mathbf{F})^*$$

has $\mathbf{u} = \mathbf{0}$ as the unique solution, then \mathbf{x}^ is an isolated zero of the VI(\mathcal{Q}, \mathbf{F}).* □

We can obtain sharpened necessary and sufficient conditions for the isolateness, and thus the globally uniqueness, of a solution to a monotone NCP(\mathbf{F}): $\mathbf{0} \leq \mathbf{x} \perp \mathbf{F}(\mathbf{x}) \geq \mathbf{0}$, where \mathbf{F} is monotone on \mathbb{R}_+^n in the usual sense, i.e.,

$$(\mathbf{x} - \mathbf{y})^T (\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})) \geq 0, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^n.$$

To state this condition, define the following index sets associated with a given solution \mathbf{x}^* of this NCP:

$$\begin{aligned} \alpha_\star &\triangleq \{j \mid x_j^* > 0 = F_j(\mathbf{x}^*)\} \quad (\text{strongly active constraints}) \\ \beta_\star &\triangleq \{j \mid x_j^* = 0 = F_j(\mathbf{x}^*)\} \quad (\text{degenerate constraints}) \\ \gamma_\star &\triangleq \{j \mid x_j^* = 0 < F_j(\mathbf{x}^*)\} \quad (\text{inactive constraints}). \end{aligned}$$

The first two parts of the result below provide necessary and sufficient conditions for \mathbf{x}^* to be an isolated solution of the monotone NCP(\mathbf{F}). In particular, part (b) of the result shows that by solving finitely many convex programs, each of which is the minimization problem of a linear objective function over a convex set, one can definitively conclude whether a given solution to a monotone NCP is isolated or not. Unlike parts (a) and (b), part (c) provides a matrix-theoretic sufficient condition for the isolatedness of the solution.

Proposition 44 *Let \mathbf{F} be a monotone function on \mathbb{R}_+^n and let \mathbf{x}^* be a solution of the NCP(\mathbf{F}). The following two statements hold:*

(a) \mathbf{x}^* is an isolated, thus unique, solution of the NCP if and only if

$$\left. \begin{array}{l} \mathbf{F}_{\alpha_*}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) = \mathbf{0} \\ \mathbf{0} \leq \mathbf{x}_{\beta_*} \perp \mathbf{F}_{\beta_*}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) \geq \mathbf{0} \end{array} \right\} \Rightarrow [\mathbf{x}_{\alpha_*} = \mathbf{x}_{\alpha_*}^*, \text{ and } \mathbf{x}_{\beta_*} = \mathbf{0}]. \quad (86)$$

(b) If \mathbf{F} is continuously differentiable near \mathbf{x}^* and $J_{\mathbf{x}_{\alpha_*}} \mathbf{F}_{\alpha_*}(\mathbf{x}^*)$ is invertible, then \mathbf{x}^* is an isolated solution of the NCP(\mathbf{F}) if and only if for every index subset $\hat{\beta} \subseteq \beta_*$ with complement $\tilde{\beta} \triangleq \beta_* \setminus \hat{\beta}$, the convex program

$$\begin{aligned} & \underset{\mathbf{x}_{\beta_*}}{\text{minimize}} \quad \sum_{k \in \hat{\beta}} x_k \\ & \text{subject to} \quad \mathbf{F}_{\alpha_*}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) = \mathbf{0} \\ & \quad \mathbf{F}_{\hat{\beta}}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) = \mathbf{0} \leq \mathbf{x}_{\hat{\beta}} \\ & \quad \text{and} \quad \mathbf{F}_{\tilde{\beta}}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) \geq \mathbf{0} = \mathbf{x}_{\tilde{\beta}} \end{aligned} \quad (87)$$

has $\mathbf{x}_{\hat{\beta}} = \mathbf{0}$ as the unique solution.

(c) If \mathbf{F} is continuously differentiable near \mathbf{x}^* and

$$\begin{bmatrix} J_{\mathbf{x}_{\alpha_*}} \mathbf{F}_{\alpha_*}(\mathbf{x}^*) & J_{\mathbf{x}_{\hat{\beta}}} \mathbf{F}_{\alpha_*}(\mathbf{x}^*) \\ J_{\mathbf{x}_{\hat{\beta}}} \mathbf{F}_{\hat{\beta}}(\mathbf{x}^*) & J_{\mathbf{x}_{\tilde{\beta}}} \mathbf{F}_{\hat{\beta}}(\mathbf{x}^*) \end{bmatrix}$$

is invertible for every index subset $\hat{\beta} \subseteq \beta_*$, then \mathbf{x}^* is an isolated solution of the NCP(\mathbf{F}).

Proof. To prove (a), suppose that x^* is the unique solution of the NCP(\mathbf{F}). Let $(\hat{\mathbf{x}}_{\alpha_*}, \hat{\mathbf{x}}_{\beta_*})$ satisfy the left-hand side of (86). Since the mapping

$$\begin{pmatrix} \mathbf{x}_{\alpha_*} \\ \mathbf{x}_{\beta_*} \end{pmatrix} \mapsto \begin{pmatrix} \mathbf{F}_{\alpha_*}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) \\ \mathbf{F}_{\beta_*}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\beta_*}, \mathbf{0}) \end{pmatrix} \quad (88)$$

remains monotone, it follows that for all $\tau > 0$ sufficiently small, $(\mathbf{x}_{\alpha_*}^* + \tau(\hat{\mathbf{x}}_{\alpha_*} - \mathbf{x}_{\alpha_*}^*), \tau \hat{\mathbf{x}}_{\beta_*}, \mathbf{0})$ is a solution of the NCP(\mathbf{F}). Hence the right-hand side of (86) holds. Conversely, suppose $\hat{\mathbf{x}}$ is an alternate solution of the NCP(\mathbf{F}). Without loss of generality, we may assume that $\hat{\mathbf{x}}$ is sufficiently near \mathbf{x}^* so that $\hat{\mathbf{x}}_{\alpha_*} > \mathbf{0}$ and $\mathbf{F}_{\gamma_*}(\hat{\mathbf{x}}) > 0$. Hence $\mathbf{F}_{\alpha_*}(\hat{\mathbf{x}}) = \mathbf{0}$ and $\hat{\mathbf{x}}_{\gamma_*} = \mathbf{0}$. By (86), we deduce that $\hat{\mathbf{x}} = \mathbf{x}^*$. This completes the proof of part (a).

For part (b), we first show that the feasible region of (87) is indeed convex. Let $(\mathbf{x}_{\alpha_*}^\ell, x_{\beta_*}^\ell)$ for $\ell = 1, 2$ be two feasible solutions to (87); these solutions must satisfy the left-hand conditions of (86). In addition to (88), the map

$$\begin{pmatrix} \mathbf{x}_{\alpha_*} \\ \mathbf{x}_{\hat{\beta}} \end{pmatrix} \mapsto \begin{pmatrix} \mathbf{F}_{\alpha_*}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\hat{\beta}}, \mathbf{0}) \\ \mathbf{F}_{\hat{\beta}}(\mathbf{x}_{\alpha_*}, \mathbf{x}_{\hat{\beta}}, \mathbf{0}) \end{pmatrix}$$

is also monotone; thus for all $\tau \in [0, 1]$, $\mathbf{z}(\tau) \triangleq \tau(\mathbf{x}_{\alpha_*}^1, \mathbf{x}_{\beta_*}^1, \mathbf{0}) + (1 - \tau)(\mathbf{x}_{\alpha_*}^2, \mathbf{x}_{\beta_*}^2, \mathbf{0})$ satisfies $\mathbf{F}_{\alpha_*}(\mathbf{z}(\tau)) = \mathbf{0}$, $\mathbf{F}_{\hat{\beta}}(\mathbf{z}(\tau)) = \mathbf{0} \leq \mathbf{z}_{\hat{\beta}}(\tau)$, and $\mathbf{F}_{\tilde{\beta}}(\mathbf{z}(\tau)) \geq \mathbf{0} = \mathbf{z}_{\tilde{\beta}}(\tau)$. This establishes the convexity of the constraint set of

(87). By the implicit function theorem, the nonsingularity $\mathbf{J}_{\mathbf{x}_{\alpha^*}} \mathbf{F}_{\alpha^*}(\mathbf{x}^*)$ implies that $\mathbf{x}_{\alpha^*}^*$ is an isolated, thus unique, zero of the equation $\mathbf{F}_{\alpha^*}(\mathbf{x}_{\alpha^*}, \mathbf{0}, \mathbf{0}) = \mathbf{0}$. It is now easy to deduce that part (b) follows from part (a). \square

B Proof of Theorem 13

It is sufficient to show that, under the assumptions of the theorem, the best-response mapping is a block-contraction (i.e., a contraction with respect to some block-maximum norm); the latter property indeed guarantees that conditions of the asynchronous convergence theorem in [53, Prop. 2.1] are satisfied by the asynchronous best-response algorithm described in Algorithm 1.

We introduce first the following norms: Given the set of nonsingular matrices \mathbf{C}_i 's coming from (23), the block-maximum norm on \mathbb{R}^n , with $n = n_1 + \dots + n_I$, is defined as

$$\|\mathbf{x}\|_{\text{block}}^{\mathbf{w}} \triangleq \max_{i=1,\dots,I} \frac{\|\mathbf{C}_i^{-1} \mathbf{x}_i\|}{c_i}, \quad \text{for } \mathbf{x} = (\mathbf{x}_i)_{i=1}^I \in \mathbb{R}^n, \quad (89)$$

where $\|\bullet\|$ is the Euclidean norm on \mathbb{R}^{n_i} and $\mathbf{c} \triangleq [c_1, \dots, c_I]^T > \mathbf{0}$ is any positive weight vector; the (weighted) maximum norm on \mathbb{R}^I is defined as (see, e.g., [53])

$$\|\mathbf{x}\|_{\infty, \text{vec}}^{\mathbf{c}} \triangleq \max_{i=1,\dots,I} \frac{|x_i|}{c_i}, \quad \text{for } \mathbf{x} \in \mathbb{R}^I; \quad (90)$$

and the matrix norm $\|\cdot\|_{\infty, \text{mat}}^{\mathbf{c}}$ on $\mathbb{R}^{I \times I}$ induced by $\|\cdot\|_{\infty, \text{vec}}^{\mathbf{c}}$, given by

$$\|\mathbf{A}\|_{\infty, \text{mat}}^{\mathbf{c}} \triangleq \max_i \frac{1}{c_i} \sum_{j=1}^I |[\mathbf{A}]_{ij}| c_j, \quad \text{for } \mathbf{A} \in \mathbb{R}^{I \times I}. \quad (91)$$

Under the under the P property of $\Upsilon_{\mathbf{F}}$, let us introduce the best-response mapping

$$\mathcal{B}(\mathbf{x}) \triangleq (\mathcal{B}_i(\mathbf{x}_{-i}))_{i=1}^I : \mathcal{Q} \ni \mathbf{x} \mapsto \mathcal{Q}, \quad \text{with} \quad \mathcal{B}_i(\mathbf{x}_{-i}) \triangleq \operatorname{argmin}_{\mathbf{x}_i \in \mathcal{Q}_i} f_i(\mathbf{x}_i, \mathbf{x}_{-i}). \quad (92)$$

The contraction properties of $\mathcal{B}(\mathbf{x})$ are given in the following, where $\mathbf{\Gamma}_{\mathbf{F}}$ is defined in (22) [note that the P property of $\Upsilon_{\mathbf{F}}$, implies the strong convexity of each $f_i(\cdot, \mathbf{x}_{-i})$ for any $\mathbf{x}_{-i} \in \mathcal{Q}_{-i}$, and thus $\alpha_i^{\min} > 0$ for all i ; hence, $\mathbf{\Gamma}_{\mathbf{F}}$ is well-defined].

Proposition 45 *Suppose that $\Upsilon_{\mathbf{F}}$ in (22) is a P matrix. Then, the best-response mapping $\mathcal{B}(\mathbf{x})$ is a block-contraction, i.e., there exists some $\mathbf{c} > \mathbf{0}$ such that*

$$\|\mathcal{B}(\mathbf{x}) - \mathcal{B}(\mathbf{y})\|_{\text{block}}^{\mathbf{c}} \leq \alpha_{\mathbf{c}} \|\mathbf{x} - \mathbf{y}\|_{\text{block}}^{\mathbf{c}} \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{Q} \quad (93)$$

with $\alpha_{\mathbf{c}} \triangleq \|\mathbf{\Gamma}_{\mathbf{F}}\|_{\infty, \text{mat}}^{\mathbf{c}} < 1$.

Proof. For any two vectors $\mathbf{x}, \mathbf{y} \in \mathcal{Q}$ we have by the minimum principle

$$\begin{aligned} (\mathbf{z}_i - \mathcal{B}_i(\mathbf{x}))^T \nabla_{\mathbf{x}_i} f_i(\mathcal{B}_i(\mathbf{x}), \mathbf{x}_{-i}) &\geq 0 & \forall \mathbf{z}_i \in \mathcal{Q}_i, \quad i = 1, \dots, I, \\ (\mathbf{z}_i - \mathcal{B}_i(\mathbf{y}))^T \nabla_{\mathbf{x}_i} f_i(\mathcal{B}_i(\mathbf{y}), \mathbf{y}_{-i}) &\geq 0 & \forall \mathbf{z}_i \in \mathcal{Q}_i, \quad i = 1, \dots, I. \end{aligned} \quad (94)$$

Substituting $\mathbf{z}_i = \mathcal{B}_i(\mathbf{y})$ into the former inequality and $\mathbf{z}_i = \mathcal{B}_i(\mathbf{x})$ into the latter, adding the two resulting inequalities we obtain with $\hat{\mathbf{z}}_i \triangleq t_i(\mathcal{B}_i(\mathbf{y}), \mathbf{y}_{-i}) + (1 - t_i)(\mathcal{B}_i(\mathbf{x}), \mathbf{x}_{-i})$ and some $t_i \in (0, 1)$:

$$\begin{aligned} 0 &\leq (\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y}))^T (\nabla_{\mathbf{x}_i} f_i(\mathcal{B}_i(\mathbf{y}), \mathbf{y}_{-i}) - \nabla_{\mathbf{x}_i} f_i(\mathcal{B}_i(\mathbf{x}), \mathbf{x}_{-i})) \\ &= (\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y}))^T \nabla_{\mathbf{x}_i \mathbf{x}_i}^2 f_i(\hat{\mathbf{z}}) (\mathcal{B}_i(\mathbf{y}) - \mathcal{B}_i(\mathbf{x})) + (\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y}))^T \sum_{j \neq i} \nabla_{\mathbf{x}_i \mathbf{x}_j}^2 f_i(\hat{\mathbf{z}}) (\mathbf{y}_j - \mathbf{x}_j) \end{aligned} \quad (95)$$

where the equality follows from the application of the main-value theorem to the univariate, differentiable, scalar function

$$[0, 1] \ni t_i \mapsto (\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y}))^T \nabla_{\mathbf{x}_i} f_i(t_i(\mathcal{B}_i(\mathbf{y}), \mathbf{y}_{-i}) + (1 - t_i)(\mathcal{B}_i(\mathbf{x}), \mathbf{x}_{-i})). \quad (96)$$

Using the definition of α_i^{\min} and β_{ij}^{\max} as given in (23), we deduce from the inequality in (95)

$$\|\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y})\|_i \alpha_i^{\min} \leq \sum_{j \neq i} \beta_{ij}^{\max} \|\mathbf{x}_j - \mathbf{y}_j\|_j, \quad i = 1, \dots, I, \quad (97)$$

(the inequality in (95) is trivially satisfied if $\|\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y})\|_i = 0$). Introducing the vectors $\mathbf{e}_{\mathcal{B}} \triangleq (e_{\mathcal{B}_i})_{i=1}^I$ and $\mathbf{e} \triangleq (e_i)_{i=1}^I$ with $e_{\mathcal{B}_i} \triangleq \|\mathcal{B}_i(\mathbf{x}) - \mathcal{B}_i(\mathbf{y})\|_i$ and $e_i = \|\mathbf{x}_i - \mathbf{y}_i\|_i$, using the definition of $\mathbf{\Gamma}_{\mathbf{F}}$ in (22), and the fact that $\alpha_i^{\min} > 0$ for all i , (97) can be written in vectorial form as

$$\mathbf{e}_{\mathcal{B}} \leq \mathbf{\Gamma}_{\mathbf{F}} \mathbf{e}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{Q}. \quad (98)$$

It follows from (98) that, for any given $\mathbf{c} > 0$, we have

$$\|\mathcal{B}(\mathbf{x}) - \mathcal{B}(\mathbf{y})\|_{\text{block}}^{\mathbf{c}} = \|\mathbf{e}_{\mathcal{B}}\|_{\infty, \text{vec}}^{\mathbf{c}} \leq \|\mathbf{\Gamma}_{\mathbf{F}}\|_{\infty, \text{mat}}^{\mathbf{c}} \|\mathbf{e}\|_{\infty, \text{vec}}^{\mathbf{c}} = \|\mathbf{\Gamma}_{\mathbf{F}}\|_{\infty, \text{mat}}^{\mathbf{c}} \|\mathbf{x} - \mathbf{y}\|_{\text{block}}^{\mathbf{c}} \quad (99)$$

which proves the inequality in (93). To complete the proof we need to show that $\|\mathbf{\Gamma}_{\mathbf{F}}\|_{\infty, \text{mat}}^{\mathbf{c}} < 1$ for some $\mathbf{c} > 0$. Invoking [51, Lemma 13.14] and [53, Cor. 6.1], we obtain the desired result:

$$\mathbf{\Upsilon}_{\mathbf{F}} \text{ is a P-matrix} \quad \Leftrightarrow \quad \exists \bar{\mathbf{c}} > 0 \quad \text{such that} \quad \|\mathbf{\Gamma}_{\mathbf{F}}\|_{\infty, \text{mat}}^{\bar{\mathbf{c}}} < 1. \quad (100)$$

■

C Proof of Proposition 17

Since \mathcal{G} is a monotone NEP, the VI associated with the NEP $\mathcal{G}_{\tau, \mathbf{y}}$ —the VI($\mathcal{Q}, \mathbf{F} + \tau(\mathbf{I} - \mathbf{y})$) with $\mathbf{F} = (\nabla_{\mathbf{x}_i} f_i)_{i=1}^I$ —is strongly monotone on \mathcal{Q} ; it follows by Theorem 8(d) that $\mathcal{G}_{\tau, \mathbf{y}}$ has a unique NE for any given $\tau > 0$ and $\mathbf{y} \in \mathbb{R}^n$. Let us denote such a unique NE by $\mathbf{S}_{\tau}(\mathbf{y}) \triangleq \text{SOL}(\mathcal{Q}, \mathbf{F} + \tau(\mathbf{I} - \mathbf{y}))$.

Necessity: Let $\mathbf{x}^* \in \mathcal{Q}$ be a NE of the monotone NEP \mathcal{G} . By Proposition 7, $\mathbf{x}^* \in \text{SOL}(\mathcal{Q}, \mathbf{F})$; then the following hold:

$$\begin{aligned} &(\mathbf{x} - \mathbf{x}^*)^T \mathbf{F}(\mathbf{x}^*) \geq 0 \quad \forall \mathbf{x} \in \mathcal{Q} \\ \Leftrightarrow &(\mathbf{x} - \mathbf{x}^*)^T (\mathbf{F}(\mathbf{x}^*) + \tau(\mathbf{x}^* - \mathbf{x}^*)) \geq 0 \quad \forall \mathbf{x} \in \mathcal{Q} \\ \Rightarrow &\mathbf{x}^* = \text{SOL}(\mathcal{Q}, \mathbf{F} + \tau(\mathbf{I} - \mathbf{x}^*)) = \mathbf{S}_{\tau}(\mathbf{x}^*), \end{aligned}$$

implying that \mathbf{x}^* is the unique NE of $\mathcal{G}_{\tau, \mathbf{x}^*}$.

Sufficiency: Let \mathbf{x}^* be a NE of $\mathcal{G}_{\tau, \mathbf{x}^*}$. Then, we have $\mathbf{x}^* = \mathbf{S}_\tau(\mathbf{x}^*)$, which leads to the desired result:

$$\begin{aligned} (\mathbf{x} - \mathbf{S}_\tau(\mathbf{x}^*))^T (\mathbf{F}(\mathbf{S}_\tau(\mathbf{x}^*)) + \tau(\mathbf{S}_\tau(\mathbf{x}^*) - \mathbf{x}^*)) &\geq 0 \quad \forall \mathbf{x} \in \mathcal{Q} \\ \Leftrightarrow (\mathbf{x} - \mathbf{x}^*)^T \mathbf{F}(\mathbf{x}^*) &\geq 0 \quad \forall \mathbf{x} \in \mathcal{Q}. \end{aligned}$$

D Proof of Theorem 22

To prove the theorem we hinge on the theory of VIs. We preliminary observe that the game \mathcal{G} is equivalent to the VI(\mathcal{Q}, \mathbf{F}), with $\mathbf{F} = (\nabla_{\mathbf{x}_i} f_i)_{i=1}^I$ (Proposition 7); SOL(\mathcal{Q}, \mathbf{f}) is thus also the solution set of the VI, i.e., SOL(\mathcal{Q}, \mathbf{F}) = SOL(\mathcal{Q}, \mathbf{f}). Moreover, still invoking Proposition 7, we have that the game $\mathcal{G}_{\tau, \varepsilon^{(n)}, \mathbf{x}^{(n)}}$ in Step 2 of Algorithm 4 is equivalent to the VI($\mathcal{Q}, \mathbf{F}^{(n)}$), where

$$\mathbf{F}^{(n)}(\mathbf{x}) \triangleq \mathbf{F}(\mathbf{x}) + \varepsilon^{(n)} \nabla \phi(\mathbf{x}) + \tau(\mathbf{x} - \mathbf{x}^{(n)}). \quad (101)$$

Observe that, under the assumptions of the theorem, $\mathbf{F}^{(n)}$ is strongly monotone [Definition 2(iii)], implying that the VI($\mathcal{Q}, \mathbf{F}^{(n)}$) has a unique solution [Theorem 3(c)] and thus $\mathbf{x}^{(n+1)}$ in Step 2 of Algorithm 4 is well defined at each iteration. Moreover, denoting by \mathcal{S} the solution set of (33), assumptions of the theorem, ensure that \mathcal{S} is nonempty, bounded, and convex. Let us introduce for each n ,

$$\delta^{(n)} \triangleq \frac{1}{2} \text{dist}(\mathbf{x}^{(n)}, \mathcal{S}) = \frac{1}{2} \|\mathbf{x}^{(n)} - P_{\mathcal{S}}(\mathbf{x}^{(n)})\|^2,$$

where $P_{\mathcal{S}}(\mathbf{y}) \triangleq \arg\min_{\mathbf{x} \in \mathcal{S}} \|\mathbf{x} - \mathbf{y}\|$ denotes the Euclidean projection on the nonempty, closed, and convex set \mathcal{S} . Then, to prove the theorem it suffices to show that the sequence $\{\delta^{(n)}\}$ converges to zero. Observe first that, since $\mathbf{x}^{(n+1)}$ at Step 2 is the solution of the game $\mathcal{G}_{\tau, \varepsilon^{(n)}, \mathbf{x}^{(n)}}$ —the VI($\mathcal{K}, \mathbf{F}^{(n)}$)—we get, for any $\mathbf{y} \in \mathcal{Q}$,

$$\left[\mathbf{F}(\mathbf{x}^{(n+1)}) + \varepsilon^{(n)} \nabla \phi(\mathbf{x}^{(n+1)}) \right]^T (\mathbf{y} - \mathbf{x}^{(n+1)}) \geq \tau (\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})^T (\mathbf{y} - \mathbf{x}^{(n+1)}). \quad (102)$$

We can write

$$\begin{aligned} \delta^{(n+1)} - \delta^{(n)} &= \frac{1}{2} \|\mathbf{x}^{(n+1)} - P_{\mathcal{S}}(\mathbf{x}^{(n+1)})\|^2 - \frac{1}{2} \|\mathbf{x}^{(n)} - P_{\mathcal{S}}(\mathbf{x}^{(n)})\|^2 \\ &\stackrel{(a)}{\leq} \frac{1}{2} \|\mathbf{x}^{(n+1)} - P_{\mathcal{S}}(\mathbf{x}^{(n)})\|^2 - \frac{1}{2} \|\mathbf{x}^{(n)} - P_{\mathcal{S}}(\mathbf{x}^{(n)})\|^2 \\ &= -\frac{1}{2} \|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2 + (\mathbf{x}^{(n)} - \mathbf{x}^{(n+1)})^T (P_{\mathcal{S}}(\mathbf{x}^{(n)}) - \mathbf{x}^{(n+1)}) \\ &\stackrel{(b)}{\leq} -\frac{1}{2} \|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2 + \frac{1}{\tau} \mathbf{F}(\mathbf{x}^{(n+1)})^T (P_{\mathcal{S}}(\mathbf{x}^{(n)}) - \mathbf{x}^{(n+1)}) \\ &\quad + \frac{\varepsilon^{(n)}}{\tau} \nabla \phi(\mathbf{x}^{(n+1)})^T (P_{\mathcal{S}}(\mathbf{x}^{(n)}) - \mathbf{x}^{(n+1)}) \\ &\stackrel{(c)}{\leq} -\frac{1}{2} \|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2 + \frac{\varepsilon^{(n)}}{\tau} \underbrace{\nabla \phi(\mathbf{x}^{(n+1)})^T (P_{\mathcal{S}}(\mathbf{x}^{(n)}) - \mathbf{x}^{(n+1)})}_{V^{(n+1)}} \end{aligned} \quad (103)$$

where: (a) follows readily from the definition of projection; (b) comes from (102) evaluated at $\mathbf{y} = P_{\mathcal{S}}(\mathbf{x}^{(n)}) \in \mathcal{Q}$; and (c) can be obtained by observing that since $P_{\mathcal{S}}(\mathbf{x}^{(n)}) \in \mathcal{S} \subseteq \text{SOL}(\mathcal{Q}, \mathbf{F})$ and $\mathbf{x}^{(n+1)} \in \mathcal{Q}$, we have $\mathbf{F}(P_{\mathcal{S}}(\mathbf{x}^{(n)}))^T (\mathbf{x}^{(n+1)} - P_{\mathcal{S}}(\mathbf{x}^{(n)})) \geq 0$, which yields in turn, by the monotonicity of \mathbf{F} [Definition 2(i)], $\mathbf{F}(\mathbf{x}^{(n+1)})^T (P_{\mathcal{S}}(\mathbf{x}^{(n)}) - \mathbf{x}^{(n+1)}) \leq 0$. We now distinguish three cases.

Case 1: Eventually, $V^{(n+1)} \leq 0$.

In this case the nonnegative sequence $\{\delta^{(n)}\}$ is (eventually) non-increasing and therefore convergent. Let us denote by n_0 the index from which all $V^{(n)}$ are non positive, and let us consider $n \geq n_0$ without loss of generality. Since \mathcal{S} is bounded, this implies that also $\{\mathbf{x}^{(n)}\}$ is bounded. Furthermore, it follows from (103) that $\{\delta^{(n+1)} - \delta^{(n)}\}$ converges to zero and $\delta^{(n+1)} - \delta^{(n)} \leq -(1/2) \|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|^2$, which shows that

$$\lim_{n \rightarrow \infty} \|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\| = 0. \quad (104)$$

Summing (103) from n_0 to $n - 1$, we get

$$\delta^{(n)} - \delta^{(n_0)} \leq \frac{1}{\tau} \sum_{i=n_0}^{n-1} \varepsilon^{(i)} V^{(i+1)}.$$

Since $\{\delta^{(n)}\}$ converges and $V^{(n)} \leq 0$, assumption (ii) in the theorem implies that $\limsup_{n \rightarrow \infty} V^{(n)} = 0$. Then, there exists a subsequence J such that

$$\lim_{\substack{n \in J \\ n \rightarrow \infty}} V^{(n)} = 0. \quad (105)$$

Since $\{\mathbf{x}^{(n)}\}$ is bounded we may assume, without loss of generality, that $\lim_{\substack{n \in J \\ n \rightarrow \infty}} \mathbf{x}^{(n)} = \tilde{\mathbf{x}}$. Note that, since \mathcal{Q} is closed, $\tilde{\mathbf{x}} \in \mathcal{Q}$. We show that actually $\tilde{\mathbf{x}} \in \text{SOL}(\mathcal{Q}, \mathbf{F})$. If this is not so, there exists a point $\mathbf{y} \in \mathcal{Q}$ such that $\mathbf{F}(\tilde{\mathbf{x}})^T (\mathbf{y} - \tilde{\mathbf{x}}) < 0$. Since $\mathbf{x}^{(n)}$ is the solution of the VI($\mathcal{K}, \mathbf{F}^{(n)}$) in Step 2 of the algorithm, we can write,

$$\left[\mathbf{F}(\mathbf{x}^{(n)}) + \tau (\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}) \right]^T (\mathbf{y} - \mathbf{x}^{(n)}) + \varepsilon^{(n-1)} \nabla \phi(\mathbf{x}^{(n)})^T (\mathbf{y} - \mathbf{x}^{(n)}) \geq 0. \quad (106)$$

By continuity, the definition of \mathbf{y} , the boundedness of $\{\mathbf{x}^{(n)}\}$, and (104), we have, without loss of generality (after a suitable renumeration),

$$\lim_{\substack{n \in J \\ n \rightarrow \infty}} \mathbf{F}(\mathbf{x}^{(n)})^T (\mathbf{y} - \mathbf{x}^{(n)}) < 0, \quad (107)$$

$$\lim_{\substack{n \in J \\ n \rightarrow \infty}} \tau (\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)})^T (\mathbf{y} - \mathbf{x}^{(n)}) = 0, \quad (108)$$

$$\lim_{\substack{n \in J \\ n \rightarrow \infty}} \varepsilon^{(n-1)} \nabla \phi(\mathbf{x}^{(n)})^T (\mathbf{y} - \mathbf{x}^{(n)}) = 0, \quad (109)$$

which contradicts (106). Therefore $\tilde{\mathbf{x}} \in \text{SOL}(\mathcal{Q}, \mathbf{F})$.

Thanks to (104) we have $\lim_{\substack{n \in J, n \rightarrow \infty}} \mathbf{x}^{(n-1)} = \tilde{\mathbf{x}}$. Therefore, by (105) and continuity, we get $\nabla \phi(\tilde{\mathbf{x}})^T (P_{\mathcal{S}}(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}) = 0$. But the convexity of ϕ implies that

$$\phi(P_{\mathcal{S}}(\tilde{\mathbf{x}})) \geq \phi(\tilde{\mathbf{x}}) + \nabla \phi(\tilde{\mathbf{x}})^T (P_{\mathcal{S}}(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}) = \phi(\tilde{\mathbf{x}}),$$

thus showing that $\tilde{\mathbf{x}} \in \mathcal{S}$. Therefore we get $\lim_{\substack{n \in J \\ n \rightarrow \infty}} \delta^{(n)} = 0$. But since the whole sequence $\{\delta^{(n)}\}$ is convergent, this implies that the entire sequence $\{\delta^{(n)}\}$ converges to 0, thus concluding the analysis of Case 1.

Case 2: The two index sets J and \bar{J} are both infinite, where $J \triangleq \{n \mid V^{(n)} > 0\}$ and

$$\bar{J} \triangleq \left\{ n \in J \mid -\frac{1}{2} \|\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}\|^2 + \frac{\varepsilon^{(n-1)}}{\tau} V^{(n)} > 0 \right\}.$$

By (103), if $n \in \bar{J}$ it might happen that $\delta^{(n)} > \delta^{(n-1)}$, while if $n \notin \bar{J}$ then necessarily $\delta^{(n)} \leq \delta^{(n-1)}$. Therefore, since \bar{J} is infinite, to prove that $\{\delta^{(n)}\}$ goes to zero it is enough to show that the subsequence $\{\delta^{(n)}\}_{\bar{J}}$ converges to zero. To this end, first observe that for every $n \in \bar{J}$ it holds that

$$\varepsilon^{(n-1)} V^{(n)} > \frac{\tau}{2} \|\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}\|^2. \quad (110)$$

The sequence $\{\mathbf{x}^{(n)}\}_{\bar{J}}$ is bounded since the definition of $V^{(n)}$, (110) and convexity imply $\phi(P_{\mathcal{S}}(\mathbf{x}^{(n-1)})) \geq \phi(\mathbf{x}^{(n)})$. But $\phi(P_{\mathcal{S}}(\mathbf{x}^{(n-1)}))$ is the optimal value of (33) and therefore is a number, say β , that does not depend on the iteration n . Therefore, since $\mathbf{x}^{(n-1)}$ belongs to \mathcal{Q} , we have that $\{\mathbf{x}^{(n)}\}_{\bar{J}}$ is bounded. By continuity, also $\{V^{(n)}\}_{\bar{J}}$ is bounded. Hence, since $\{\varepsilon^{(n)}\}$ converges to 0, (110) yields

$$\lim_{\substack{n \in \bar{J} \\ n \rightarrow \infty}} \|\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}\| = 0. \quad (111)$$

Since $\{\mathbf{x}^{(n)}\}_{\bar{J}}$ is bounded, it has limit points. Let $\tilde{J} \subseteq \bar{J}$ be a subsequence such that $\lim_{\substack{n \in \tilde{J} \\ n \rightarrow \infty}} \mathbf{x}^{(n)} = \tilde{\mathbf{x}}$. Reasoning exactly as in Case 1, (the only difference is that instead of (104) we use (111)), we may deduce that $\tilde{\mathbf{x}} \in \text{SOL}(\mathcal{Q}, \mathbf{F})$. By continuity, $\nabla \phi(\tilde{\mathbf{x}})^T (P_{\mathcal{S}}(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}) \geq 0$. Thus $\tilde{\mathbf{x}} \in \mathcal{S}$; hence $\lim_{\substack{n \in \tilde{J} \\ n \rightarrow \infty}} \delta^{(n)} = 0$. Since this reasoning can be repeated for every convergent subsequence of $\{\mathbf{x}^{(n)}\}_{\bar{J}}$, we conclude that $\lim_{\substack{n \in \bar{J} \\ n \rightarrow \infty}} \delta^{(n)} = 0$, thus concluding the analysis of this case.

Case 3: The index set J is infinite while \bar{J} is finite. In this case, the sequence $\{\delta^{(n)}\}$ is non-increasing eventually. Therefore $\{\delta^{(n)}\}$ converges, implying that $\{\mathbf{x}^{(n)}\}$ is bounded, $\{\delta^{(n+1)} - \delta^{(n)}\}$ converges to zero and therefore, by (103), also (104) holds. At this point, we can proceed exactly as in Case 1 and Case 2 to prove that $\{\delta^{(n)}\}$ converges to zero, thus concluding the proof of the theorem. \square

E Proof of Lemma 24

To prove the lemma it is sufficient to show that the first-order Taylor expansion as given in (42) holds for the function f ; the rest of the proof follows similar steps as those used to prove the minimum principle for (real-valued) functions of real variables and thus is omitted; see e.g., [58].

Before proving the lemma, it is useful to introduce a real-coordinate representation of real-valued functions of complex matrices and establish the connection between standard derivatives of this representation and the \mathbb{R} -derivatives of the original functions of complex variables.

The complex space $\mathbb{C}^{n \times m}$ of dimension $n \cdot m$ has a natural structure of a real space \mathbb{R}^{2nm} of dimensions $2n \cdot m$; this comes readily, e.g., from the following isomorphic transformation:

$$\mathbb{C}^{n \times m} \ni \mathbf{Z} \iff \tilde{\mathbf{z}} \triangleq \begin{bmatrix} \text{vec}(\text{Re}(\mathbf{Z})) \\ \text{vec}(\text{Im}(\mathbf{Z})) \end{bmatrix} \in \mathbb{R}^{2nm}. \quad (112)$$

For the sake of simplicity, in the following, we will denote by $\mathcal{Z} \triangleq \mathbb{C}^{n \times m}$ the original complex space and by \mathbf{Z} the elements of \mathcal{Z} ; $\mathcal{R} \triangleq \mathbb{R}^{2nm}$ will be the $2n \cdot m$ -dimensional space of real vectors in the form $\check{\mathbf{z}}$, i.e.,

$$\mathcal{R} \triangleq \left\{ \check{\mathbf{z}} \in \mathbb{R}^{2nm} : \check{\mathbf{z}} \triangleq \begin{bmatrix} \check{\mathbf{z}}_R \\ \check{\mathbf{z}}_I \end{bmatrix} \triangleq \begin{bmatrix} \text{vec}(\text{Re}(\mathbf{Z})) \\ \text{vec}(\text{Im}(\mathbf{Z})) \end{bmatrix}, \text{ for some } \mathbf{Z} \in \mathcal{Z} \right\}; \quad (113)$$

elements of \mathcal{R} will be denoted by $\check{\mathbf{z}}$, and partitioned as in (113).

Given a real-valued function of complex matrices $f : \mathcal{Z} \rightarrow \mathbb{R}$, the representation of $f(\mathbf{Z})$ under the isomorphic transformation (112) is denoted by $\check{f}(\check{\mathbf{z}}) = f(\mathbf{Z})$. Note that if $f(\mathbf{Z})$ is \mathbb{R} -(continuously) differentiable on \mathcal{Z} then $\check{f}(\check{\mathbf{z}})$ is (continuously) differentiable on \mathcal{R} . Moreover, we can easily establish the connection between the Jacobian of $\check{f}(\check{\mathbf{z}})$ and the Jacobian and conjugate Jacobian of $f(\mathbf{Z})$ [cf. (39)], as shown next. By definition, for any $\check{\mathbf{z}} \in \mathcal{R}$, the Jacobian of $\check{f}(\check{\mathbf{z}})$ is

$$D_{\check{\mathbf{z}}} \check{f}(\check{\mathbf{z}}) \triangleq (\nabla_{\check{\mathbf{z}}} \check{f}(\check{\mathbf{z}}))^T = \left[\frac{\partial \check{f}(\check{\mathbf{z}})}{\partial \check{\mathbf{z}}_R^T}, \frac{\partial \check{f}(\check{\mathbf{z}})}{\partial \check{\mathbf{z}}_I^T} \right] \triangleq [D_{\check{\mathbf{z}}_R} \check{f}(\check{\mathbf{z}}), D_{\check{\mathbf{z}}_I} \check{f}(\check{\mathbf{z}})]. \quad (114)$$

Using (37) and (39), it is not difficult to see that, for any given $\mathcal{Z} \ni \mathbf{Z} \iff \check{\mathbf{z}} \in \mathcal{R}$, the following hold

$$D_{\mathbf{Z}} f(\mathbf{Z}) \triangleq \frac{\partial f(\mathbf{Z})}{\partial \text{vec}(\mathbf{Z})^T} = \frac{1}{2} [D_{\check{\mathbf{z}}_R} \check{f}(\check{\mathbf{z}}) - j \cdot D_{\check{\mathbf{z}}_I} \check{f}(\check{\mathbf{z}})], \quad (115)$$

$$D_{\mathbf{Z}^*} f(\mathbf{Z}) \triangleq \frac{\partial f(\mathbf{Z})}{\partial \text{vec}(\mathbf{Z}^*)^T} = \frac{1}{2} [D_{\check{\mathbf{z}}_R} \check{f}(\check{\mathbf{z}}) + j \cdot D_{\check{\mathbf{z}}_I} \check{f}(\check{\mathbf{z}})], \quad (116)$$

which provides the desired relationship between $D_{\check{\mathbf{z}}} \check{f}(\check{\mathbf{z}})$ and $D_{\mathbf{Z}} f(\mathbf{Z})$ and $D_{\mathbf{Z}^*} f(\mathbf{Z})$.

Exploring the above equivalences, we can now readily prove Lemma 24 leveraging on standard real calculus results. Given a real-valued convex and continuously \mathbb{R} -differentiable function $f : \mathcal{K} \rightarrow \mathbb{R}$ on \mathcal{K} , the first-order Taylor expansion of $f(\mathbf{Z}) = \check{f}(\check{\mathbf{z}})$ at $\mathcal{K} \ni \mathbf{Z}_0 (\iff \check{\mathbf{z}}_0)$ exists and it is given by:

$$\begin{aligned} f(\mathbf{Z}_0 + \Delta \mathbf{Z}) - f(\mathbf{Z}_0) &= \check{f}(\check{\mathbf{z}}_0 + \Delta \check{\mathbf{z}}) - \check{f}(\check{\mathbf{z}}_0) \\ &\simeq D_{\check{\mathbf{z}}} \check{f}(\check{\mathbf{z}}_0) \cdot \Delta \check{\mathbf{z}} \end{aligned} \quad (117)$$

$$= \text{Re} \left\{ [D_{\check{\mathbf{z}}_R} \check{f}(\check{\mathbf{z}}_0) - j \cdot D_{\check{\mathbf{z}}_I} \check{f}(\check{\mathbf{z}}_0)] [(\Delta \check{\mathbf{z}})_R + j \cdot (\Delta \check{\mathbf{z}})_I] \right\} \quad (118)$$

$$= 2 \text{Re} \left\{ D_{\mathbf{Z}} f(\mathbf{Z}_0) \text{vec}(\Delta \mathbf{Z}) \right\} \quad (119)$$

$$= 2 \text{Re} \left\{ \text{tr} \left((\nabla_{\mathbf{Z}} f(\mathbf{Z}_0))^T \Delta \mathbf{Z} \right) \right\} \quad (120)$$

$$= 2 \text{Re} \left\{ \text{tr} \left((\nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0))^H \Delta \mathbf{Z} \right) \right\} \quad (121)$$

$$= 2 \langle \Delta \mathbf{Z}, \nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0) \rangle \quad (122)$$

where (117) follows from the first-order Taylor expansion of real-valued functions of real vectors (see, e.g., [58]); (119) follows from (115); (120) is due to $D_{\mathbf{Z}} f(\mathbf{Z}_0) = \text{vec}(\nabla_{\mathbf{Z}} f(\mathbf{Z}_0))^T$ and the property $\text{vec}(\mathbf{A})^T \text{vec}(\mathbf{B}) = \text{tr}(\mathbf{A}^T \mathbf{B})$ for any $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n \times m}$; and in (121) we used the fact that f is real and thus $(\nabla_{\mathbf{Z}} f(\mathbf{Z}_0))^* = \nabla_{\mathbf{Z}^*} f(\mathbf{Z}_0)$. This completes the proof. \square

F Complex Matrix Derivatives in Example 25

We derive here the expressions of the (conjugate) derivatives used in the Example 25 and Example 25 revisited.

In order to obtain the expression of the augmented Hessian $\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*}\tilde{f}(\mathbf{Z})$, we need to compute $\nabla_{\mathbf{Z}\mathbf{Z}^*}^2\tilde{f}(\mathbf{Z}) \triangleq D_{\mathbf{Z}}\left(\nabla_{\mathbf{Z}^*}\tilde{f}(\mathbf{Z})\right)$ and $\nabla_{\mathbf{Z}^*\mathbf{Z}^*}^2\tilde{f}(\mathbf{Z}) \triangleq D_{\mathbf{Z}^*}\left(\nabla_{\mathbf{Z}^*}\tilde{f}(\mathbf{Z})\right)$. We preliminary compute $\nabla_{\mathbf{Z}}f(\mathbf{Z})$ and $\nabla_{\mathbf{Z}^*}f(\mathbf{Z})$.

Recalling that [40, Prop. 3.12] $d\ln\det(\mathbf{Z}) = \text{Tr}(\mathbf{Z}^{-1}d\mathbf{Z})$ for all \mathbf{Z} such that $\det\mathbf{Z} \neq 0$, with $d\ln\det(\mathbf{Z})$ being the (complex) differential of $\ln\det(\mathbf{Z})$, we have (up to a constant positive factor)

$$df(\mathbf{Z}) = \text{Tr}\left((\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}d\mathbf{Z}\mathbf{H}^H\right) = \text{vec}^T\left(\left(\mathbf{H}^H(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}\right)^T\right)\text{vec}(d\mathbf{Z}) \quad (123)$$

which, using the identification rule as given in [40, Table 3.3], leads to the following Jacobian matrices of f :

$$D_{\mathbf{Z}}f(\mathbf{Z}) = \text{vec}^T\left(\left(\mathbf{H}^H(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}\right)^T\right) \quad \text{and} \quad D_{\mathbf{Z}^*}f(\mathbf{Z}) = \mathbf{0}, \quad (124)$$

and thus [cf. (39)]

$$\nabla_{\mathbf{Z}}f(\mathbf{Z}) = \left(\mathbf{H}^H(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}\right)^T \quad \text{and} \quad \nabla_{\mathbf{Z}^*}f(\mathbf{Z}) = \mathbf{0}. \quad (125)$$

It follows from (125) that

$$\nabla_{\mathbf{Z}^*}\tilde{f}(\mathbf{Z}) = \nabla_{\mathbf{Z}^*}f(\mathbf{Z}) + \nabla_{\mathbf{Z}^*}f(\mathbf{Z})^* = (\nabla_{\mathbf{Z}}f(\mathbf{Z}))^* = \mathbf{H}^H(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}. \quad (126)$$

Given (126), we can now compute $\nabla_{\mathbf{Z}\mathbf{Z}^*}^2\tilde{f}(\mathbf{Z})$ and $\nabla_{\mathbf{Z}^*\mathbf{Z}^*}^2\tilde{f}(\mathbf{Z})$. The differential of $\nabla_{\mathbf{Z}^*}\tilde{f}(\mathbf{Z})$ is:

$$\text{vec}\left[d\left(\nabla_{\mathbf{Z}^*}\tilde{f}(\mathbf{Z}^*)\right)\right] = \text{vec}\left[\mathbf{H}^H d(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}\right] \quad (127)$$

$$= - \underbrace{\text{vec}\left[\mathbf{G}(\mathbf{Z})(d\mathbf{Z})^H\mathbf{G}(\mathbf{Z})\right]}_{\mathbf{G}(\mathbf{Z}) \triangleq \mathbf{H}^H(\mathbf{R}_n + \mathbf{H}\mathbf{Z}\mathbf{H}^H)^{-1}\mathbf{H}} \quad (128)$$

$$= -[\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})] \text{vec}[(d\mathbf{Z}^*)^T] \quad (129)$$

$$= -[\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})] \mathbf{K}_{n^2n^2} \text{vec}[(d\mathbf{Z}^*)] \quad (130)$$

where in (128) we used the rule $d\mathbf{Z}^{-1} = -\mathbf{Z}^{-1}(d\mathbf{Z})\mathbf{Z}^{-1}$ [40, Prop. 3.8]; (129) follows from the property $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ [40, Lemma 2.11]; and in the last equality we introduced the commutation matrix $\mathbf{K}_{n^2n^2}$, which is the $n^2 \times n^2$ permutation matrix such that $\text{vec}(\mathbf{A}^T) = \mathbf{K}_{n^2n^2}\text{vec}(\mathbf{A})$ [40, Def. 1.8].

It follows from (130) and the identification rule [40, Table 3.3] that

$$\nabla_{\mathbf{Z}^*\mathbf{Z}^*}^2\tilde{f}(\mathbf{Z}) = -[\mathbf{G}(\mathbf{Z})^T \otimes \mathbf{G}(\mathbf{Z})] \mathbf{K}_{n^2n^2}, \quad (131)$$

$$\nabla_{\mathbf{Z}\mathbf{Z}^*}^2\tilde{f}(\mathbf{Z}) = \mathbf{0}, \quad (132)$$

which leads to the expression of the augmented Hessian $\mathcal{H}_{\mathbf{Z}\mathbf{Z}^*}\tilde{f}(\mathbf{Z})$ as given in (54).

G Proof of Propositions 28 and 29

It is sufficient to prove only Proposition 29; Proposition 28 is just a special case. To do that, we need the following intermediate result.

G.1 Mean-value theorem for functions of complex variables

We provide here a version of the mean-value theorem that is suitable for real-valued functions of complex matrices. We focus directly on the specific function that we need to prove Proposition 29.

Given a continuously \mathbb{R} -differentiable matrix function $\mathbf{F}^{\mathbb{C}} : \mathcal{K} \rightarrow \mathbb{C}^{n \times m}$ on the convex and closed set $\mathcal{K} \subseteq \mathbb{C}^{n \times m}$ and a point $\Delta \mathbf{Y} \in \mathbb{C}^{n \times m}$, let us consider the real-valued function of complex matrix variables

$$g(\mathbf{Z}) \triangleq \langle \Delta \mathbf{Y}, \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \rangle. \quad (133)$$

For every two points $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathcal{K}$, with $\Delta \mathbf{Z} \triangleq \mathbf{Z}_2 - \mathbf{Z}_1$, let $h(t) : [0, 1] \rightarrow \mathbb{R}$ be the real-valued scalar function, defined as: with $\mathbf{Z}(t) \triangleq \mathbf{Z}_1 + t \Delta \mathbf{Z}$,

$$[0, 1] \ni t \mapsto h(t) \triangleq \langle \Delta \mathbf{Y}, \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t)) \rangle. \quad (134)$$

For some $\bar{t} \in (0, 1)$, we have

$$g(\mathbf{Z}_2) - g(\mathbf{Z}_1) = h(1) - h(0) = h'(\bar{t}) \quad (135)$$

where $h'(t)$ is the first order derivative of $h(t)$ [note that h is continuously differentiable on $(0, 1)$], and the last equality in (135) follows from the mean-value theorem applied to the function $h(t)$. To compute $h'(t)$ we use the chain rule for complex matrix derivatives [40] as shown next. Rewriting $h(t)$ as

$$h(t) = \frac{1}{2} \text{tr} \left(\Delta \mathbf{Y}^H \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t)) \right) + \frac{1}{2} \text{tr} \left(\Delta \mathbf{Y}^T \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t))^* \right) \quad (136)$$

and using

$$\begin{aligned} D_{\mathbf{Z}} \text{tr} \left(\Delta \mathbf{Y}^H \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \right) &= \text{vec}(\Delta \mathbf{Y}^*)^T D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \\ D_{\mathbf{Z}^*} \text{tr} \left(\Delta \mathbf{Y}^H \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \right) &= \text{vec}(\Delta \mathbf{Y}^*)^T D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \\ D_{\mathbf{Z}} \text{tr} \left(\Delta \mathbf{Y}^T \mathbf{F}^{\mathbb{C}}(\mathbf{Z})^* \right) &= \left(D_{\mathbf{Z}^*} \text{tr} \left(\Delta \mathbf{Y}^H \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \right) \right)^* \\ D_{\mathbf{Z}^*} \text{tr} \left(\Delta \mathbf{Y}^T \mathbf{F}^{\mathbb{C}}(\mathbf{Z})^* \right) &= \left(D_{\mathbf{Z}} \text{tr} \left(\Delta \mathbf{Y}^H \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \right) \right)^* \end{aligned}$$

we have

$$\begin{aligned} h'(t) &= D_{\mathbf{Z}(t)} h(t) D_t \mathbf{Z}(t) + D_{\mathbf{Z}(t)^*} h(t) D_t \mathbf{Z}(t)^* \\ &= \frac{1}{2} \text{vec}(\Delta \mathbf{Y}^*)^T D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t)) \text{vec}(\Delta \mathbf{Z}) + \frac{1}{2} \text{vec}(\Delta \mathbf{Y})^T D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t))^* \text{vec}(\Delta \mathbf{Z}) \\ &\quad + \frac{1}{2} \text{vec}(\Delta \mathbf{Y}^*)^T D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t)) \text{vec}(\Delta \mathbf{Z}^*) + \frac{1}{2} \text{vec}(\Delta \mathbf{Y})^T D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t))^* \text{vec}(\Delta \mathbf{Z}^*) \\ &= \frac{1}{2} \text{vec}([\Delta \mathbf{Y}, \Delta \mathbf{Y}^*])^H \begin{bmatrix} D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t)) & D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t)) \\ D_{\mathbf{Z}} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t))^* & D_{\mathbf{Z}^*} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(t))^* \end{bmatrix} \text{vec}([\Delta \mathbf{Z}, \Delta \mathbf{Z}^*]), \end{aligned} \quad (137)$$

where in (137) we used the chain rule [40, Th. 3.1]. Using (138) and the augmented Jacobian matrix $\mathbf{JF}^{\mathbb{C}}(\mathbf{Z})$ introduced in (46), we can rewrite (135) in a compact form as

$$g(\mathbf{Z}_2) - g(\mathbf{Z}_1) = \text{vec}([\Delta \mathbf{Y}, \Delta \mathbf{Y}^*])^H \mathbf{JF}^{\mathbb{C}}(\mathbf{Z}(\bar{t})) \text{vec}([\Delta \mathbf{Z}, \Delta \mathbf{Z}^*]). \quad (139)$$

which is the desired result.

G.2 Proof of Proposition 29

We prove only (a)-(c); the proof of (d)-(e) follows similar steps.

Sufficiency part. For (a)-(c), it is enough to prove only (c). Given two points $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathcal{K}$, let us define $\Delta \mathbf{Z} \triangleq \mathbf{Z}_2 - \mathbf{Z}_1$; we have, for some $\bar{t} \in (0, 1)$,

$$\left\langle \mathbf{Z}_2 - \mathbf{Z}_1, \mathbf{F}^{\mathbb{C}}(\mathbf{Z}_2) - \mathbf{F}^{\mathbb{C}}(\mathbf{Z}_1) \right\rangle = \text{vec}([\Delta \mathbf{Z}, \Delta \mathbf{Z}^*])^H \mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}(\bar{t})) \text{vec}([\Delta \mathbf{Z}, \Delta \mathbf{Z}^*]), \quad (140)$$

where the equality follows from (139). Since $\mathbf{Z}_1, \mathbf{Z}_2 \in \mathcal{K}$, we have that $\Delta \mathbf{Z} \in S_{\mathcal{K}}$; moreover $\mathbf{Z}(\bar{t}) \in \mathcal{K}$ (due to the convexity of \mathcal{K}). It follows from (140) that if there exists a constant c_{sm} such that $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq (c_{\text{sm}}/2) \|\mathbf{Y}\|_F^2$ for all $\mathbf{Y} \in S_{\mathcal{K}}$ and $\mathbf{Z} \in \mathcal{K}$, then $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ is strongly monotone on \mathcal{K} .

Necessity part. Let us focus on the strongly monotonicity property only; monotonicity is obtained in a similar way. Suppose that $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ is strongly monotone on \mathcal{K} with constant $c_{\text{sm}} > 0$. Let us show first that $\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq (c_{\text{sm}}/2) \|\mathbf{Y}\|_F^2$ for all $\mathbf{Y} \in S_{\mathcal{K}}$ and $\mathbf{Z} \in \text{ri}(\mathcal{K})$, where $\text{ri}(\mathcal{K})$ denotes the relative interior of \mathcal{K} (see [58, Ch. 1.4] for the definition of $\text{ri}(\mathcal{K})$ and its main properties). Then, we have

$$\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) = \frac{1}{2} \text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \lim_{t \downarrow 0} \frac{1}{t} \begin{bmatrix} \text{vec}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z} + t\mathbf{Y}) - \mathbf{F}^{\mathbb{C}}(\mathbf{Z})) \\ \text{vec}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z} + t\mathbf{Y}) - \mathbf{F}^{\mathbb{C}}(\mathbf{Z}))^* \end{bmatrix} \quad (141)$$

$$= \frac{1}{2} \lim_{t \downarrow 0} \frac{1}{t^2} \left\langle t\mathbf{Y}, \mathbf{F}^{\mathbb{C}}(\mathbf{Z} + t\mathbf{Y}) - \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \right\rangle \quad (142)$$

$$\geq \frac{c_{\text{sm}}}{2} \lim_{t \downarrow 0} \frac{1}{t^2} \|t\mathbf{Y}\|_F^2 = \frac{c_{\text{sm}}}{2} \|\mathbf{Y}\|_F^2, \quad \forall \mathbf{Y} \in S_{\mathcal{K}} \text{ and } \mathbf{Z} \in \text{ri}(\mathcal{K}), \quad (143)$$

where the equality in (141) follows from the (\mathbb{R} -)differentiability of $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ on \mathcal{K} [(141) can be proved using the same approach as in the proof of Lemma 24 but applied to $\text{vec}(\mathbf{F}^{\mathbb{C}}(\mathbf{Z}))$]; in (142) we used the definition of inner product (43); and in (143) we used i) the fact that for sufficiently small $t > 0$, $\mathbf{Z} + t\mathbf{Y} \in \mathcal{K}$ [since $\mathbf{Z} \in \text{ri}(\mathcal{K})$], and ii) the strongly monotonicity of $\mathbf{F}^{\mathbb{C}}(\mathbf{Z})$ on \mathcal{K} .

Next, let $\mathbf{Z} \in \mathcal{K}$ but $\mathbf{Z} \notin \text{ri}(\mathcal{K})$; by [58, Proposition 1.4.1(a)] there exists a sequence $\{\mathbf{Z}_k\} \subset \text{ri}(\mathcal{K})$ such that $\mathbf{Z}_k \rightarrow \mathbf{Z}$. By (143) evaluated in each $\mathbf{Z}_k \in \text{ri}(\mathcal{K})$ and the continuity of $\mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z})$, it follows that

$$\text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) = \lim_{k \rightarrow \infty} \text{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathbf{J} \mathbf{F}^{\mathbb{C}}(\mathbf{Z}_k) \text{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq \frac{c_{\text{sm}}}{2} \|\mathbf{Y}\|_F^2,$$

for all $\mathbf{Y} \in S_{\mathcal{K}}$. This completes the proof of the necessity part.

Note that Proposition 29 reduces to Proposition 28 if the set \mathcal{K} has nonempty interior. Indeed, when this happens, $\text{Aff}(\mathcal{K}) = \mathbb{C}^{n \times m}$ and thus $S_{\mathcal{K}} = \mathbb{C}^{n \times m}$.

H Proof of Lemma 35 and Lemma 36

H.1 Proof of Lemma 35

Consider the quadratic equation $\tilde{\mu}_k + \tau(p_k - c_k) = \frac{H_k}{1 + H_k p_k}$ or, equivalently,

$$\tau H_k p_k^2 + (\tau + H_k(\tilde{\mu}_k - \tau c_k)) p_k + (\tilde{\mu}_k - \tau c_k - H_k) = 0.$$

We can write the equation as $ap_k^2 + bp_k + c = 0$ with the obvious identifications.

- i) To show that both roots $(-b \pm \sqrt{b^2 - 4ac})/2a$ are real, it suffices to show that $b^2 \geq 4ac$. After several manipulations, it follows that

$$b^2 - 4ac = [\tau - H_k(\tilde{\mu}_k - \tau c_k)]^2 + 4\tau H_k^2 \geq 0.$$

- ii) We will show the decreasingness through the derivative:

$$\begin{aligned} \frac{d}{d\tilde{\mu}_k} (-b \pm \sqrt{b^2 - 4ac}) &= -H_k \pm \frac{H_k(\tau - H_k(\tilde{\mu}_k - \tau c_k))}{\sqrt{b^2 - 4ac}} \\ (\text{up to a nonnegative scaling factor}) &= -H_k \sqrt{b^2 - 4ac} \pm H_k(\tau - H_k(\tilde{\mu}_k - \tau c_k)) \\ &= -H_k \left(\sqrt{[\tau - H_k(\tilde{\mu}_k - \tau c_k)]^2 + 4\tau H_k^2} \pm (\tau - H_k(\tilde{\mu}_k - \tau c_k)) \right) \\ &< -H_k \left(\sqrt{[\tau - H_k(\tilde{\mu}_k - \tau c_k)]^2} \pm (\tau - H_k(\tilde{\mu}_k - \tau c_k)) \right) \\ &= -H_k(|\tau - H_k(\tilde{\mu}_k - \tau c_k)| \pm (\tau - H_k(\tilde{\mu}_k - \tau c_k))) \\ &\leq 0. \end{aligned}$$

- iii) It suffices to show that $b + \sqrt{b^2 - 4ac}$ is positive for $\tilde{\mu}_k = 0$ combined with the decreasingness. Indeed, for $\tilde{\mu}_k = 0$, $b = \tau - H_k \tau c_k$ and

$$\begin{aligned} b^2 - 4ac &= [\tau + H_k \tau c_k]^2 + 4\tau H_k^2 \\ &= [\tau - H_k \tau c_k]^2 + 4\tau H_k(\tau c_k + H_k). \end{aligned}$$

It follows that

$$\begin{aligned} b + \sqrt{b^2 - 4ac} &= (\tau - H_k \tau c_k) + \sqrt{[\tau - H_k \tau c_k]^2 + 4\tau H_k(\tau c_k + H_k)} \\ &> (\tau - H_k \tau c_k) + |\tau - H_k \tau c_k| \\ &\geq 0 \end{aligned}$$

where we have used $c_k \geq 0$.

- iv) The expression for the positive root follows from $(-b + \sqrt{b^2 - 4ac})/2a$ together with $b^2 - 4ac = [\tau - H_k(\tilde{\mu}_k - \tau c_k)]^2 + 4\tau H_k^2$, $b = \tau + H_k(\tilde{\mu}_k - \tau c_k)$, and $a = \tau H_k$.

H.2 Proof of Lemma 36

To simplify the exposition, we will absorb the fixed parameters $\boldsymbol{\mu}_{1:i-1}$ in λ_k as $\lambda_k \leftarrow \lambda_k + \boldsymbol{\mu}_{1:i-1}^T \mathbf{w}_{k,1:i-1}$, we will denote μ_i and w_{ki} simply as μ and v_k , and we will denote $\boldsymbol{\mu}_{i+1:N_c}$ and $\mathbf{w}_{k,i+1:N_c}$ as $\boldsymbol{\rho}$ and \mathbf{w}_k . Therefore, instead of writing $\lambda_k + \boldsymbol{\mu}^T \mathbf{w}_k$ we will more explicitly and conveniently write $\lambda_k + \mu v_k + \boldsymbol{\rho}^T \mathbf{w}_k$. With this new notation, the goal of the proof is to show the decreasingness of $\sum_{k=1}^N v_k p_k^*(\mu)$ in $\mu \geq 0$.

First of all, observe that, for fixed parameters $\boldsymbol{\mu}_{1:i-1}$, $\mathbf{p}^*(\mu)$ is unique and continuous on $[0, \infty)$, as shown next. Given μ , the vector $\mathbf{p}^*(\mu)$ defined in (71) can be interpreted as a solution of the VI($\mathcal{K}, \nabla_{\mathbf{p}} \mathcal{L}(\cdot, \boldsymbol{\rho}^*) + \mu \mathbf{v}$), where $\mathcal{K} \triangleq \{\mathbf{p} \in \mathbb{R}^N : \mathbf{0} \leq \mathbf{p} \leq \mathbf{p}^{\max}\}$ and $\mathcal{L}(\cdot, \boldsymbol{\rho}) \triangleq -\sum_{k=1}^N [\log(1 + H_k p_k) - (\lambda_k + \boldsymbol{\rho}^T \mathbf{w}_k) p_k] + \frac{\tau}{2} \|\mathbf{p} - \mathbf{c}\|^2$,

with $\boldsymbol{\rho}^*$ satisfying the complementarity conditions $\mathbf{0} \leq \boldsymbol{\rho}^* \perp \sum_{k=1}^N \mathbf{w}_k p_k^*(\mu) \geq \mathbf{0}$. It is not difficult to see that $\nabla_{\mathbf{p}} \mathcal{L}(\cdot, \boldsymbol{\rho})$ (and thus $\nabla_{\mathbf{p}} \mathcal{L}(\cdot, \boldsymbol{\rho}^*) + \mu \mathbf{v}$) is strongly monotone on \mathcal{K} , for any given $\boldsymbol{\rho}$. It follows that: i) the solution $\mathbf{p}^*(\mu)$ of the VI($\mathcal{K}, \nabla_{\mathbf{p}} \mathcal{L}(\cdot, \boldsymbol{\rho}^*) + \mu \mathbf{v}$) is unique, and ii) $\mathbf{p}^*(-\mu)$ is a co-coercive function of μ [30, Prop. 2.3.11] implying that $\mathbf{p}^*(\mu)$ is Lipschitz continuous.

Now, the optimal solution $\mathbf{p}^*(\mu)$ in (72) contains three cases depending on whether $p_k^* = 0$, $p_k^* = p_k^{\max}$, or $p_k^* \in (0, p_k^{\max})$. The parameter μ naturally partitions $[0, \infty)$ into intervals where one of the three cases is active for each k . It suffices to assume that μ lies in the interior of one of these intervals and show then the decreasingness of $\sum_{k=1}^N v_k p_k^*(\mu)$ in the interior of each interval. This, together with the continuity of $\sum_{k=1}^N v_k p_k^*(\mu)$, implies the decreasingness over the whole interval $[0, \infty)$.

Depending on the value of μ , some of the inequalities in $\sum_{k=1}^N \mathbf{w}_k p_k^* \leq \boldsymbol{\alpha}$ will be active and some inactive. The inactive ones will have the corresponding elements in $\boldsymbol{\rho}$ equal to zero. From now on, we will only consider the active ones and, for simplicity of notation, we will assume that only the active ones are included in the vectors \mathbf{w}_k and $\boldsymbol{\alpha}$ so that we can effectively write $\sum_{k=1}^N \mathbf{w}_k p_k^* = \boldsymbol{\alpha}$. (Note that if no constraint is active, then $\sum_{k=1}^N v_k p_k^*(\mu)$ is trivially decreasing in μ since each p_k^* is decreasing in μ .)

Since, by assumption, μ is in the interior of an interval where the three cases of the power expression in (72) do not change for each k , it follows that in a sufficiently small neighborhood of μ the indexes k where either $p_k^* = 0$ or $p_k^* = p_k^{\max}$ do not change and, therefore, do not affect any of the terms in $\sum_{k=1}^N \mathbf{w}_k p_k^* = \boldsymbol{\alpha}$. We can then ignore those indexes and only consider the ones where $p_k^* \in (0, p_k^{\max})$, i.e.,

$$p_k^* = \frac{1}{2} \left(c_k - \frac{1}{H_k} \right) - \frac{1}{2\tau} \left[\tilde{\mu}_k - \sqrt{\left[\tilde{\mu}_k - \tau \left(c_k + \frac{1}{H_k} \right) \right]^2 + 4\tau} \right]$$

where $\tilde{\mu}_k = \lambda_k + \mu v_k + \boldsymbol{\rho}^T \mathbf{w}_k$. At this point, we can write the system of equations as $\sum_k \mathbf{w}_k p_k^* = \tilde{\boldsymbol{\alpha}}$, where the summation is over the indexes for which $p_k^* \in (0, p_k^{\max})$. There is yet another simplification we can make: if the vectors \mathbf{w}_k are not linearly independent, we can just use another set of vectors with reduced dimension represent the same system of linear equations. With that reduction in mind, we will still use the same notation and write $\sum_k \mathbf{w}_k p_k^* = \tilde{\boldsymbol{\alpha}}$.

The active constraints can then be rewritten as the following system of equations:

$$\mathbf{f}(\mu, \boldsymbol{\rho}) \triangleq \sum_k \mathbf{w}_k \left(\frac{1}{2} \left(c_k - \frac{1}{H_k} \right) - \frac{1}{2\tau} \left[\tilde{\mu}_k - \sqrt{\left[\tilde{\mu}_k - \tau \left(c_k + \frac{1}{H_k} \right) \right]^2 + 4\tau} \right] \right) - \tilde{\boldsymbol{\alpha}} = \mathbf{0}$$

where the summation is only over the indexes that satisfy $p_k^* \in (0, p_k^{\max})$ and $\tilde{\boldsymbol{\alpha}}$ contains $\boldsymbol{\alpha}$ minus the effect of the indexes that we are not considering in the summation (i.e., those for which either $p_k^* = 0$ or $p_k^* = p_k^{\max}$).

At this point, we can invoke the implicit function theorem to characterize $\boldsymbol{\rho}(\mu)$. In particular, the partial derivatives of $\mathbf{f}(\mu, \boldsymbol{\rho})$ are

$$\begin{aligned} \frac{\partial \mathbf{f}}{\partial \mu} &= \sum_k \mathbf{w}_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \frac{\partial \tilde{\mu}_k}{\partial \mu} = \sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \mathbf{w}_k v_k \\ \frac{\partial \mathbf{f}}{\partial \boldsymbol{\rho}^T} &= \sum_k \mathbf{w}_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \frac{\partial \tilde{\mu}_k}{\partial \boldsymbol{\rho}^T} = \sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \mathbf{w}_k \mathbf{w}_k^T \end{aligned}$$

and then

$$\frac{d\boldsymbol{\rho}(\mu)}{d\mu} = - \left(\sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \mathbf{w}_k \mathbf{w}_k^T \right)^{-1} \left(\sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \mathbf{w}_k v_k \right)$$

where the inverse exists because of the assumption of \mathbf{w}_k being linearly independent and $\partial p_k^* / \partial \tilde{\mu}_k < 0$ for all k such that $p_k^* \in (0, p_k^{\max})$ (Lemma 35).

We can now write the derivative of the weighted sum power as

$$\begin{aligned} \frac{d}{d\mu} \left(\sum_k v_k p_k^* \right) &= \frac{\partial}{\partial \mu} \left(\sum_k v_k p_k^* \right) + \frac{\partial}{\partial \boldsymbol{\rho}^T} \left(\sum_k v_k p_k^* \right) \frac{d\boldsymbol{\rho}(\mu)}{d\mu} \\ &= \sum_k v_k^2 \frac{\partial p_k^*}{\partial \tilde{\mu}_k} - \left(\sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} v_k \mathbf{w}_k^T \right) \left(\sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \mathbf{w}_k \mathbf{w}_k^T \right)^{-1} \left(\sum_k \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \mathbf{w}_k v_k \right). \end{aligned}$$

Proving that this derivative is nonpositive is equivalent to showing (since $\partial p_k^* / \partial \mu_k < 0$)

$$\left(\sum_k \left| \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \right| v_k \mathbf{w}_k^T \right) \left(\sum_k \left| \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \right| \mathbf{w}_k \mathbf{w}_k^T \right)^{-1} \left(\sum_k \left| \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \right| \mathbf{w}_k v_k \right) \leq \sum_k v_k^2 \left| \frac{\partial p_k^*}{\partial \tilde{\mu}_k} \right| \quad (144)$$

which follows from Cauchy-Schwartz's inequality for random vectors: let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ be two random vectors such that $E[\|\mathbf{x}\|^2] < \infty$, $E[\|\mathbf{y}\|^2] < \infty$ and $E[\mathbf{y}\mathbf{y}^T]$ is nonsingular, then $E[\mathbf{x}\mathbf{y}^T] E^{-1}[\mathbf{y}\mathbf{y}^T] E[\mathbf{y}\mathbf{x}^T] \leq E[\mathbf{x}\mathbf{x}^T]$. Note that the inequality in (144) is strict because of the linear independence of \mathbf{w}_k 's. \square

I Proof of Proposition 38

Statement (a) follows from Proposition 30; for (b) and (c), we prove only (b).

According to Proposition 28(a), we need to show that, under the assumption that $\boldsymbol{\Upsilon}_{\mathbf{F}^{\mathbb{C}}}^{\text{mimo}}$ in (74) is positive semidefinite, the augmented Jacobian matrix $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$ associated to $\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$ in (73) is augmented positive semidefinite on $\mathcal{P}^{\text{mimo}}$. Given (73), $D_{\mathbf{Q}}\mathbf{F}^{\mathbb{C}}(\mathbf{Q}) = \mathbf{0}$, implying that $\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$ is a block diagonal matrix:

$$\mathbf{J}\mathbf{F}^{\mathbb{C}}(\mathbf{Q}) = \frac{1}{2} \begin{bmatrix} D_{\mathbf{Q}}\mathbf{F}^{\mathbb{C}}(\mathbf{Q}) & \mathbf{0} \\ \mathbf{0} & (D_{\mathbf{Q}}\mathbf{F}^{\mathbb{C}}(\mathbf{Q}))^* \end{bmatrix} \quad (145)$$

with $D_{\mathbf{Q}}\mathbf{F}^{\mathbb{C}}(\mathbf{Q})$ given by (see Appendix F for a similar computation):

$$D_{\mathbf{Q}}\mathbf{F}^{\mathbb{C}}(\mathbf{Q}) = \begin{bmatrix} D_{\mathbf{Q}_1}\mathbf{F}_1^{\mathbb{C}}(\mathbf{Q},) & \cdots & D_{\mathbf{Q}_I}\mathbf{F}_1^{\mathbb{C}}(\mathbf{Q}) \\ \vdots & \ddots & \vdots \\ D_{\mathbf{Q}_1}\mathbf{F}_I^{\mathbb{C}}(\mathbf{Q}) & \cdots & D_{\mathbf{Q}_Q}\mathbf{F}_I^{\mathbb{C}}(\mathbf{Q}) \end{bmatrix} \quad (146)$$

$$= \begin{bmatrix} \Psi_{11}(\mathbf{Q})^* \otimes \Psi_{11}(\mathbf{Q}) & \cdots & \Psi_{1I}(\mathbf{Q})^* \otimes \Psi_{1I}(\mathbf{Q}) \\ \vdots & \ddots & \vdots \\ \Psi_{I1}(\mathbf{Q})^* \otimes \Psi_{I1}(\mathbf{Q}) & \cdots & \Psi_{II}(\mathbf{Q})^* \otimes \Psi_{II}(\mathbf{Q}) \end{bmatrix} \quad (147)$$

with

$$\Psi_{ij}(\mathbf{Q}) \triangleq \mathbf{H}_{ii}^H \underbrace{\left(\mathbf{R}_{n_i} + \sum_{j=1}^Q \mathbf{H}_{ij} \mathbf{Q}_j \mathbf{H}_{ij}^H \right)^{-1}}_{\triangleq \mathbf{S}_i(\mathbf{Q})} \mathbf{H}_{ij} = \mathbf{H}_{ii}^H \mathbf{S}_i(\mathbf{Q}) \mathbf{H}_{ij}. \quad (148)$$

We will denote by $\Psi_{ii}^{1/2}(\mathbf{Q})$ the square root of the positive definite matrix $\Psi_{ii}(\mathbf{Q})$ (recall that the channel matrices \mathbf{H}_{ii} are assumed to be full-column rank), i.e., $\Psi_{ii}(\mathbf{Q}) = \Psi_{ii}^{H/2}(\mathbf{Q}) \Psi_{ii}^{1/2}(\mathbf{Q})$.

Therefore, $\mathbf{JF}^{\mathbf{C}}(\mathbf{Q})$ is augmented positive semidefinite on $\mathcal{P}^{\text{mimo}}$ if and only if $D_{\mathbf{Q}}\mathbf{F}^{\mathbf{C}}(\mathbf{Q})$ is positive semidefinite on $\mathcal{P}^{\text{mimo}}$, or equivalently the following matrix is so:

$$\begin{bmatrix} \mathbf{I} & \cdots & \left(\Psi_{11}^{-H/2} \Psi_{1I} \Psi_{11}^{-1/2}\right)^* \otimes \left(\Psi_{11}^{-H/2} \Psi_{1I} \Psi_{11}^{-1/2}\right) \\ \vdots & \ddots & \vdots \\ \left(\Psi_{II}^{-H/2} \Psi_{I1} \Psi_{II}^{-1/2}\right)^* \otimes \left(\Psi_{II}^{-H/2} \Psi_{I1} \Psi_{II}^{-1/2}\right) & \cdots & \mathbf{I} \end{bmatrix} \quad (149)$$

where for notational simplicity we omitted the dependence on \mathbf{Q} and write Ψ_{ij} , instead of $\Psi_{ij}(\mathbf{Q})$. The condensed matrix associated to (149) is the following $I \times I$ matrix

$$\begin{bmatrix} 1 & \cdots & \left\| \left(\Psi_{11}^{-H/2} \Psi_{1Q} \Psi_{11}^{-1/2}\right)^* \otimes \left(\Psi_{11}^{-H/2} \Psi_{1Q} \Psi_{11}^{-1/2}\right) \right\|_2 \\ \vdots & \ddots & \vdots \\ \left\| \left(\Psi_{II}^{-H/2} \Psi_{I1} \Psi_{II}^{-1/2}\right)^* \otimes \left(\Psi_{II}^{-H/2} \Psi_{I1} \Psi_{II}^{-1/2}\right) \right\|_2 & \cdots & 1 \end{bmatrix}, \quad (150)$$

where $\|\mathbf{A}\|_2 \triangleq \sqrt{\rho(\mathbf{A}^H \mathbf{A})}$ is the spectral norm of \mathbf{A} . Note that we can rewrite each of the off-diagonal terms of (150) as: with $\tilde{\Psi}_{ij} \triangleq \Psi_{ii}^{-H/2} \Psi_{ij} \Psi_{ii}^{-1/2}$,

$$\begin{aligned} \left\| \left(\Psi_{ii}^{-H/2} \Psi_{ij} \Psi_{ii}^{-1/2}\right)^* \otimes \left(\Psi_{ii}^{-H/2} \Psi_{ij} \Psi_{ii}^{-1/2}\right) \right\|_2 &= \left\| \tilde{\Psi}_{ij}^* \otimes \tilde{\Psi}_{ij} \right\|_2 = \left[\rho \left(\tilde{\Psi}_{ij}^T \tilde{\Psi}_{ij}^* \otimes \tilde{\Psi}_{ij}^H \tilde{\Psi}_{ij} \right) \right]^{1/2} \\ &= \rho \left(\tilde{\Psi}_{ij}^H \tilde{\Psi}_{ij} \right), \end{aligned} \quad (151)$$

where in the last equality we used the property $\rho(\mathbf{A}^T \mathbf{A}^* \otimes \mathbf{A}^H \mathbf{A}) = \rho(\mathbf{A}^T \mathbf{A}^*) \rho(\mathbf{A}^H \mathbf{A})$ and the fact that the eigenvalues of $\mathbf{A}^T \mathbf{A}^*$ coincide with those of $\mathbf{A}^H \mathbf{A}$. Using (151), we can now introduce the so-called comparison matrix $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q})$ associated to (150) and defined as

$$[\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q})]_{ij} \triangleq \begin{cases} 1, & \text{if } i = j \\ -\rho \left(\tilde{\Psi}_{ij}^H(\mathbf{Q}) \tilde{\Psi}_{ij}(\mathbf{Q}) \right) & \text{otherwise.} \end{cases}$$

It is indeed not difficult to see that if $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q})$ is positive semidefinite on $\mathcal{P}^{\text{mimo}}$ then so is the matrix (150) and thus also $D_{\mathbf{Q}}\mathbf{F}^{\mathbf{C}}(\mathbf{Q})$. To conclude the proof, it is enough to show that $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q}) \geq \Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}$ for all $\mathbf{Q} \in \mathcal{P}^{\text{mimo}}$, where $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}$ is defined in (74) and the inequality has to be intended component-wise. The latter properties indeed implies that if $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}$ is positive semidefinite then so is $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q})$ on $\mathcal{P}^{\text{mimo}}$. To this end, we focus next on the off-diagonal terms $\rho(\tilde{\Psi}_{ij}^H(\mathbf{Q}) \tilde{\Psi}_{ij}(\mathbf{Q}))$ of $\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q})$ and prove that $|[\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}(\mathbf{Q})]_{ij}| \leq |[\Upsilon_{\mathbf{F}^{\mathbf{C}}}^{\text{mimo}}]_{ij}|$ for all $\mathbf{Q} \in \mathcal{P}^{\text{mimo}}$ and $i \neq j$. Denoting by $\mathbf{S}_i^{1/2}(\mathbf{Q})$ the square root of the positive definite matrix $\mathbf{S}_i(\mathbf{Q})$ defined in (148) [i.e., $\mathbf{S}_i(\mathbf{Q}) = \mathbf{S}_i^{H/2}(\mathbf{Q}) \mathbf{S}_i^{1/2}(\mathbf{Q})$], and using $\tilde{\Psi}_{ij}(\mathbf{Q}) \triangleq \Psi_{ii}^{-H/2}(\mathbf{Q}) \Psi_{ij}(\mathbf{Q}) \Psi_{ii}^{-1/2}(\mathbf{Q})$, we have

the following chain of equalities/inequalities: for all $\mathbf{Q} \in \mathcal{P}^{\text{mimo}}$ and $i \neq j$,

$$\begin{aligned}
|[\boldsymbol{\Upsilon}_{\mathbf{F}^{\text{C}}}^{\text{mimo}}(\mathbf{Q})]_{ij}| &= \rho\left(\tilde{\boldsymbol{\Psi}}_{ij}^H(\mathbf{Q})\tilde{\boldsymbol{\Psi}}_{ij}(\mathbf{Q})\right) = \rho\left(\boldsymbol{\Psi}_{ij}^H(\mathbf{Q})\boldsymbol{\Psi}_{ii}^{-1}(\mathbf{Q})\boldsymbol{\Psi}_{ij}(\mathbf{Q})\boldsymbol{\Psi}_{ii}^{-1}(\mathbf{Q})\right) \\
&= \rho\left(\boldsymbol{\Psi}_{ii}^{-H/2}(\mathbf{Q})\mathbf{H}_{ij}^H\mathbf{S}_i(\mathbf{Q})\mathbf{H}_{ii}\left(\mathbf{H}_{ii}^H\mathbf{S}_i(\mathbf{Q})\mathbf{H}_{ii}\right)^{-1}\mathbf{H}_{ii}^H\mathbf{S}_i(\mathbf{Q})\mathbf{H}_{ij}\boldsymbol{\Psi}_{ii}^{-1/2}(\mathbf{Q})\right) \\
&= \rho\left(\boldsymbol{\Psi}_{ii}^{-H/2}(\mathbf{Q})\mathbf{H}_{ij}^H\mathbf{S}_i^{H/2}(\mathbf{Q})\underbrace{\mathbf{S}_i^{1/2}(\mathbf{Q})\mathbf{H}_{ii}\left(\mathbf{H}_{ii}^H\mathbf{S}_i(\mathbf{Q})\mathbf{H}_{ii}\right)^{-1}\mathbf{H}_{ii}^H\mathbf{S}_i^{H/2}(\mathbf{Q})}_{\triangleq \mathbf{P}_{\mathcal{R}(\mathbf{H}_{ii})} \preceq \mathbf{I}}\mathbf{S}_i^{1/2}(\mathbf{Q})\mathbf{H}_{ij}\boldsymbol{\Psi}_{ii}^{-1/2}(\mathbf{Q})\right)
\end{aligned} \tag{152}$$

$$\leq \rho\left(\mathbf{S}_i^{1/2}(\mathbf{Q})\mathbf{H}_{ij}\boldsymbol{\Psi}_{ii}^{-1}(\mathbf{Q})\mathbf{H}_{ij}^H\mathbf{S}_i^{H/2}(\mathbf{Q})\right) = \rho\left(\mathbf{S}_i^{1/2}(\mathbf{Q})\mathbf{H}_{ij}\left(\mathbf{H}_{ii}^H\mathbf{S}_i(\mathbf{Q})\mathbf{H}_{ii}\right)^{-1}\mathbf{H}_{ij}^H\mathbf{S}_i^{H/2}(\mathbf{Q})\right) \tag{153}$$

$$\leq \rho\left(\mathbf{S}_i^{1/2}(\mathbf{Q})\mathbf{H}_{ij}\mathbf{H}_{ii}^\dagger\mathbf{S}_i^{-1}(\mathbf{Q})\mathbf{H}_{ii}^\dagger\mathbf{H}_{ij}^H\mathbf{S}_i^{H/2}(\mathbf{Q})\right) \tag{154}$$

$$\leq \rho\left(\mathbf{S}_i^{-1}(\mathbf{Q})\right) \cdot \rho\left(\mathbf{S}_i(\mathbf{Q})\right) \cdot \rho\left(\mathbf{H}_{ii}^\dagger\mathbf{H}_{ij}^H\mathbf{H}_{ij}\mathbf{H}_{ii}^\dagger\right) \tag{155}$$

$$\leq \text{INNR}_{ij} \cdot \rho\left(\mathbf{H}_{ii}^\dagger\mathbf{H}_{ij}^H\mathbf{H}_{ij}\mathbf{H}_{ii}^\dagger\right) = |[\boldsymbol{\Upsilon}_{\mathbf{F}^{\text{C}}}^{\text{mimo}}]_{ij}| \tag{156}$$

where in (152), $\mathbf{P}_{\mathcal{R}(\mathbf{H}_{ii})} \triangleq \mathbf{S}_i^{1/2}(\mathbf{Q})\mathbf{H}_{ii}\left(\mathbf{H}_{ii}^H\mathbf{S}_i(\mathbf{Q})\mathbf{H}_{ii}\right)^{-1}\mathbf{H}_{ii}^H\mathbf{S}_i^{H/2}(\mathbf{Q})$ is the orthogonal projection onto the range space of \mathbf{H}_{ii} ; in (153) we used the property of the projection $\mathbf{P}_{\mathcal{R}(\mathbf{H}_{ii})} \preceq \mathbf{I}$ and the spectral radius inequality $\rho(\mathbf{A}^H\mathbf{B}\mathbf{A}) \leq \rho(\mathbf{A}^H\mathbf{C}\mathbf{A})$ for all $\mathbf{0} \preceq \mathbf{B} \preceq \mathbf{C}$; (154) follows from the property $(\mathbf{X}^H\mathbf{A}\mathbf{X})^{-1} \preceq \mathbf{X}^\dagger\mathbf{A}^{-1}\mathbf{X}^{\dagger H}$, valid for all positive definite $n \times n$ matrices \mathbf{A} and $n \times k$ full-column rank matrices \mathbf{X} ; in (155) we used $\mathbf{A} \preceq \rho(\mathbf{A}) \cdot \mathbf{I}$ and the spectral radius inequality as in (153); and finally (156) follows from $\rho(\mathbf{S}_i^{-1}(\mathbf{Q})) \cdot \rho(\mathbf{S}_i(\mathbf{Q})) \leq \text{INNR}_{ij}$, with INNR_{ij} defined in (75).

The above chain of inequalities proves the desired relationship between $\boldsymbol{\Upsilon}_{\mathbf{F}^{\text{C}}}^{\text{mimo}}(\mathbf{Q})$ and $\boldsymbol{\Upsilon}_{\mathbf{F}^{\text{C}}}^{\text{mimo}}$, which completes the proof. \square

References

- [1] W. Yu, G. Ginis, and J. M. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE Jour. on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, June 2002.
- [2] Z.-Q. Luo and J.-S. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP Jour. on Applied Signal Processing*, vol. 2006, pp. 1–10, May 2006.
- [3] K. W. Shum, K.-K. Leung, and C. W. Sung, "Convergence of iterative waterfilling algorithm for gaussian interference channels," *IEEE J. Selected Areas in Comm.*, vol. 25, no. 6, pp. 1091–1100, Aug. 2007.
- [4] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory—part I&II: Nash equilibria and distributed algorithms," *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1230–1267, March 2008.
- [5] —, "Asynchronous iterative water-filling for Gaussian frequency-selective interference channels," *IEEE Trans. on Information Theory*, vol. 54, no. 7, pp. 2868–2878, July 2008.
- [6] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang, "Distributed power allocation with rate constraints in Gaussian parallel interference channels," *IEEE Trans. on Information Theory*, vol. 54, no. 8, pp. 3471–3489, Aug. 2008.
- [7] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: A unified view," *IEEE Jour. on Selected Areas in Communications*, vol. 26, no. 7, pp. 1089–1103, Sept. 2008.

- [8] —, “The MIMO iterative waterfilling algorithm,” *IEEE Trans. on Signal Processing*, vol. 57, no. 5, May 2009.
- [9] —, “Cognitive MIMO radio: Competitive optimality design based on subspace projections,” *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 46–59, Nov. 2008.
- [10] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, “Flexible design of cognitive radio wireless systems: From game theory to variational inequality theory,” *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 107–123, September 2009.
- [11] G. Scutari and D. P. Palomar, “MIMO cognitive radio: A game theoretical approach,” *IEEE Trans. on Signal Processing*, vol. 58, no. 2, pp. 761–780, February 2010.
- [12] M. Razaviyayn, Z.-Q. Luo, P. Tseng, and J.-S. Pang, “A stackelberg game approach to distributed spectrum management,” *Mathematical Programming, Series B, Special Issue on "Large Scale Optimization: Analysis, Algorithms and Applications"*, vol. 129, no. 2.
- [13] E. Altman and L. Wynter, “Equilibrium, games, and pricing in transportation and telecommunications networks,” *Networks and Spatial Economics*, vol. 4, no. 1, pp. 7–21, March 2004.
- [14] E. Altman, T. Boulogne, R. E. Azouzi, T. Jimenez, and L. Wynter, “A survey on networking games,” *Computers and Operations Research*, vol. 33, pp. 286–311, 2006.
- [15] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, “Autonomous spectrum balancing for digital subscriber lines,” *IEEE Trans. on Signal Processing*, vol. 55, no. 8, pp. 4241–4257, Aug. 2007.
- [16] W. Saad, Z. Han, H. Poor, and T. Basar, “Game theoretic methods for the smart grid,” *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, September 2012.
- [17] I. Atzeni, L. G. Ordonez, G. Scutari, and D. P. Palomar, “Demand-side management via distributed energy generation and storage optimization,” *IEEE Trans. on Smart Grid*, vol. 28, (to appear) 2012.
- [18] J. Huang, D. P. Palomar, N. Mandayam, J. Walrand, S. Wicker, and T. Basar, eds., (Special Issue on Game Theory in Communication Systems), *IEEE J. Selected Areas in Comm.*, vol. 26, no. 7, Sept 2008.
- [19] E. Jorswieck, E. G. Larsson, M. Luise, , and V. H. Poor, eds., (Special Issue on Game Theory in Signal Processing and Communications), *IEEE Signal Processing Mag.*, vol. 26, no. 5, Sept 2009.
- [20] F. Facchinei and J.-S. Pang, “Nash equilibria: The variational approach,” in *Convex Optimization in Signal Processing and Communications*, D. P. Palomar and Y. C. Eldar, Eds. London: Cambridge University Press, 2009, ch. 12.
- [21] F. Facchinei, J.-S. Pang, G. Scutari, and L. Lampariello, “Vi-constrained hemivariational inequalities: Distributed algorithms and power control in ad-hoc networks,” *Mathematical Programming (submitted)*, 2011.
- [22] F. Facchinei, V. Piccialli, and M. Sciandrone, “Decomposition algorithms for generalized potential games,” *Computational Optimization and Applications*, vol. 50, no. 2, pp. 237–262, 2011.
- [23] J. Huang, R. Berry, and M. L. Honig, “Distributed interference compensation for wireless networks,” *IEEE Jour. on Selected Areas in Communications*, vol. 24, no. 5, pp. 1074–1084, May 2006.
- [24] G. Arslan, M. F. Demirkol, and Y. Song, “Equilibrium efficiency improvement in MIMO interference systems: a decentralized stream control approach,” *IEEE Trans. on Wireless Communications*, vol. 6, no. 8, pp. 2984–2993, Aug. 2007.
- [25] J. Rosen, “Existence and uniqueness of equilibrium points for concave n-person games,” *Econometrica*, vol. 33, no. 3, pp. 520–534, July 1965.
- [26] T. Basar and G. J. Olsder, *Dynamic noncooperative game theory*. 2nd ed. Academic Press, London, 1989.
- [27] S. Li and T. Basar, “Distributed algorithms for the computation of noncooperative equilibria,” *Automatica*, vol. 23, no. 4, pp. 523–533, June 1987.
- [28] S. Uryasev and R. Y. Rubinstein, “On relaxation algorithms in computation of noncooperative equilibria,” *IEEE Trans. on Automatic Control*, vol. 39, no. 6, pp. 1263–1267, 1994.

- [29] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Jour. on Selected Areas in Communications*, vol. 52, no. 7, pp. 1341–1347, Sept. 1995.
- [30] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problem*. Springer-Verlag, New York, 2003.
- [31] J.-S. Pang, G. Scutari, D. P. Palomar, and F. Facchinei, "Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, pp. 3251–3271, June 2010.
- [32] G. Scutari, D. Palomar, F. Facchinei, and J.-S. Pang, "Convex optimization, game theory, and variational inequality theory in multiuser communication systems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 35–49, May 2010.
- [33] I. V. Konnov, *Equilibrium Models and Variational Inequalities*. Elsevier B.V., Amsterdam, 2007.
- [34] H. Yin, U. Shanbhag, and P. G. Mehta, "Nash equilibrium problems with congestion costs and shared constraints," *IEEE Trans. on Automatic Control*, vol. 56, no. 7, pp. 1702–1708, July 2011.
- [35] A. Kannan and U. Shanbhag, "Distributed computation of equilibria in monotone nash games via iterative regularization techniques," *SIAM Journal of Optimization*, (under review). [Online]. Available: [https://netfiles.uiuc.edu/udaybag/www/Papers/siopt\\$_\\$iter\\$_\\$_\\$final\\$_\\$_\\$submitted.pdf](https://netfiles.uiuc.edu/udaybag/www/Papers/siopt$_$iter$_$_$final$_$_$submitted.pdf)
- [36] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to nash equilibria," *IEEE Trans. on Automatic Control*, vol. 50, no. 3, pp. 312–327, March 2005.
- [37] J.-S. Pang, "Asymmetric variational inequality problems over product sets: Applications and iterative methods," *Mathematical Programming*, vol. 31, no. 2, pp. 206–219, June 1985.
- [38] J.-S. Pang and D. Chan, "Iterative methods for variational and complementarity problems," *Mathematical Programming*, vol. 24, no. 1, pp. 284–313, Dec. 1982.
- [39] G. Scutari, F. Facchinei, J.-S. Pang, and L. Lampariello, "Equilibrium selection in power control games on the interference channel," in *Proc. of the 31st Annual IEEE International Conference on Computer Communications (IEEE INFOCOM 2012)*, Orlando, Florida, USA, Mach 25-30 2012.
- [40] A. Hjørungnes, *Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*. London: Cambridge University Press, May 2011.
- [41] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Jour. on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [42] Z.-Q. Luo and S. Zhang, "Spectrum management: Complexity and duality," *Jour. of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–72, Feb. 2008.
- [43] R. Cendrillon, J. Huang, M. Chiang, and M. Moonen, "Autonomous spectrum balancing for digital subscriber lines," *IEEE Trans. on Signal Processing*, vol. 55, no. 8, pp. 4241–4257, Aug. 2007.
- [44] R. Zhang, Y. C. Liang, and S. Cui, "Dynamic resource allocation in cognitive radio networks," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 102–114, May 2010.
- [45] E. Larsson and E. Jorswieck, "Competition versus collaboration on the MISO interference channel," *IEEE Jour. on Selected Areas in Communications*, vol. 26, no. 7, pp. 1059–1069, Sept. 2008.
- [46] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communication Magazine*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [47] I. Guvenc, S. Saunders, O. Oyman, H. Claussen, and A. Gatherer, "Special issue on femtocell networks," *EURASIP Journal on Wireless Communications and Networking*, vol. Volume 2010, Article ID 367878, 2 pages, 2010. doi:10.1155/2010/367878.
- [48] V. Chandrasekhar and L. Andrews, "Uplink capacity and interference avoidance for two-tier femtocell networks," *IEEE Trans. on Wireless Communications*, vol. 7, no. 8, pp. 1536–1276, July 2009.

- [49] S. Barbarossa, A. Carfagna, S. Sardellitti, M. Olimpo, and L. Pescosolido, "Optimal radio access in femtocell networks based on Markov modeling of interferers' activity," in *Proc. of the IEEE 2011 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 22-27, 2011.
- [50] P. Billingsley, "Statistical methods in markov chains," *The Annals of Mathematical Statistics*, vol. 32, no. 1, pp. 12–40, March 1961.
- [51] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*. SIAM Classics in Applied Mathematics 60, Philadelphia (2009) [Originally published by Academic Press, Boston (1992)].
- [52] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial Mathematics (SIAM), 1987.
- [53] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Athena Scientific Press, 1989.
- [54] R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications*. Cambridge University Press, 1997.
- [55] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial Mathematics (SIAM), Phila, USA, 1987.
- [56] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. on Communications*, vol. 50, no. 2, pp. 291–303, Feb. 2002.
- [57] K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," Electrical and Computer Engineering, Jacobs School of Engineering, University of California, San Diego, Tech. Rep., July 2009.
- [58] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Scientific, 2003.
- [59] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, "Distributed dynamic pricing for mimo interfering multiuser systems: A unified approach," in *Proc. of International Conference on NETwork Games, COntrol and OPTimization (NetGCoP 2011)*, Paris, France, October 2011, pp. 12–14.
- [60] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for mimo ad hoc cognitive radio networks," *IEEE Trans. on Information Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.